

QY	61	GTGATGAGACGTGTCCCACTGAGTGGCCACAGCAGAGTGTGAGCATGGGCTGAG	120
Db			
QY	61	GTGATGAGACGTGTCCCACTGAGTGGCCACAGCAGAGTGTGAGCATGGGCTGAG	120
Db			
QY	121	AGCTGGACCGGACCAAGGCTGGCGAGAAATGGGCGCTGGCTGATTCCTAGGACGTT	180
Db			
QY	121	AGCTGGACCGGACCAAGGCTGGCGAGAAATGGGCGCTGGCTGATTCCTAGGACGTT	180
Db			
QY	181	GGCGGACAGAGGAGGAGCGGCGAGCTCTGGAGCAGAGCCGAGAGCAGTCTG	240
Db			
QY	181	GGCGGACAGAGGAGGAGCGGCGAGCTCTGGAGCAGAGCCGAGAGCAGTCTG	240
Db			
QY	241	GAGTGCCTGAAGCGGCGGCTGAGCCCTACCGGCTGGCCCACTATGTTCCAGAGGCTGTG	300
Db			
QY	241	GAGTGCCTGAAGCGGCGGCTGAGCCCTACCGGCTGGCCCACTATGTTCCAGAGGCTGTG	300
Db			
QY	301	GGTGGCGGCTGTGGCGGACCGGAAAGCCAGCTCTTGTGTGCTAACCTGTAACTTT	360
Db			
QY	301	GGTGGCGGCTGTGGCGGACCGGAAAGCCAGCTCTTGTGTGCTAACCTGTAACTTT	360
Db			
QY	361	TGGCGTGGAGGTGCTTTGGCGGAGGAGCATCACCTATGTGCGGCTCTGTGCTGGAAGT	420
Db			
QY	361	TGGCGTGGAGGTGCTTTGGCGGAGGAGCATCACCTATGTGCGGCTCTGTGCTGGAAGT	420
Db			
QY	421	GGGGGTAGAGGAGAAATTTCATGACCATGGTGTGGGCAATGGTCCAGTGTGGGCTGGT	480
Db			
QY	421	GGGGGTAGAGGAGAAATTTCATGACCATGGTGTGGGCAATGGTCCAGTGTGGGCTGGT	480
Db			
QY	481	CTGTGTCGCGCTCTAGGCTAGCCAGTACCATGGGCTGAGGCTATGGCGGCGGCGG	540
Db			
QY	481	CTGTGTCGCGCTCTAGGCTAGCCAGTACCATGGGCTGAGGCTATGGCGGCGGCGG	540
Db			
QY	541	GCCCTTCATCTGGGCACTGTCTTGGGCACTCTGTGAGGCTCTTTCATCCCAAGGC	600
Db			
QY	541	GCCCTTCATCTGGGCACTGTCTTGGGCACTCTGTGAGGCTCTTTCATCCCAAGGC	600
Db			
QY	601	CGGCTGGCTAGCAGGCTGTGTGCCGAGTCCAGGCGGCTGGAGCTGGGCTGCTCAT	660
Db			
QY	601	CGGCTGGCTAGCAGGCTGTGTGCCGAGTCCAGGCGGCTGGAGCTGGGCTGCTCAT	660
Db			
QY	661	CGTGGCGTGGGCTGTGGCTGTGGGCACTCTGTGAGGCTGTCTTCTTCTTCTTCTTCTT	720
Db			
QY	661	CGTGGCGTGGGCTGTGGCTGTGGGCACTCTGTGAGGCTGTCTTCTTCTTCTTCTTCTT	720
Db			
QY	721	GCTCTGTGACCTCTTCCGGGACCGGACCACTGTGCCAGGCTACTGTCTATGCTT	780
Db			
QY	721	GCTCTGTGACCTCTTCCGGGACCGGACCACTGTGCCAGGCTACTGTCTATGCTT	780
Db			
QY	781	CATGATCAGTCTTGGGGCTGCTGGGCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTG	840
Db			
QY	781	CATGATCAGTCTTGGGGCTGCTGGGCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTG	840
Db			
QY	841	TGCGCTGGGCGGCTACCTGGGCACTGGGAGGAGTCTTCTTGGGCTGCTGCTGCTGCTG	900
Db			
QY	841	TGCGCTGGGCGGCTACCTGGGCACTGGGAGGAGTCTTCTTGGGCTGCTGCTGCTGCTG	900
Db			
QY	901	CTTCTCTACCTCGTAGCAGCCACACTGCTGTGGCTGAGGAGGAGGCTGGGCGGCGC	960
Db			
QY	901	CTTCTCTACCTCGTAGCAGCCACACTGCTGTGGCTGAGGAGGAGGCTGGGCGGCGC	960
Db			
QY	961	CGAGCCAGCAGAGGCTGTGGGCGGCTGCTTGTGCGGCGGCTGCTGCTGCTGCTGCTG	1020
Db			
QY	961	CGAGCCAGCAGAGGCTGTGGGCGGCTGCTTGTGCGGCGGCTGCTGCTGCTGCTGCTG	1020
Db			
QY	1021	CGGCTGGCTTTCGGGAACTGGGCGGCTGCTTCCCGGCTGACAGGCTGTGTGCTG	1080
Db			
QY	1021	CGGCTGGCTTTCGGGAACTGGGCGGCTGCTTCCCGGCTGACAGGCTGTGTGCTG	1080
Db			
QY	1081	CATGCCCCGACCTGTGCGGCGGCTCTTCTGTGCTGAGCTGTGAGCTGTGAGTGTG	1140
Db			
QY	1081	CATGCCCCGACCTGTGCGGCGGCTCTTCTGTGCTGAGCTGTGAGCTGTGAGTGTG	1140
Db			

QY	1141	GACCTTACGCTGCTTTTACACGATTTCTGTGGCGAGGGCTGTACACAGGGGCTGCCAG	1200
Db			
QY	1141	GACCTTACGCTGCTTTTACACGATTTCTGTGGCGAGGGCTGTACACAGGGGCTGCCAG	1200
Db			
QY	1201	AGCTGAGCGGGCACAGGCGGAGACATATGATGAAGGCTTTCGATGGGAGGCT	1260
Db			
QY	1201	AGCTGAGCGGGCACAGGCGGAGACATATGATGAAGGCTTTCGATGGGAGGCT	1260
Db			
QY	1261	GGGCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1320
Db			
QY	1261	GGGCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1320
Db			
QY	1321	GCAGCGATTCGCGACCTCGAGCAGTATTTGGCCAGTGTGGAGCTTCCCTGTGGCTGC	1380
Db			
QY	1321	GCAGCGATTCGCGACCTCGAGCAGTATTTGGCCAGTGTGGAGCTTCCCTGTGGCTGC	1380
Db			
QY	1381	CGGTGCCACATGCTCTGCCACAGTGTGGCGGCTGACAGCTTACAGCGGCTCACCGG	1440
Db			
QY	1381	CGGTGCCACATGCTCTGCCACAGTGTGGCGGCTGACAGCTTACAGCGGCTCACCGG	1440
Db			
QY	1441	GTTACACCTTCTAGCCCTCGAGATCTGCCCTACACACTGGGCTTCCCTTACACCGGA	1500
Db			
QY	1441	GTTACACCTTCTAGCCCTCGAGATCTGCCCTACACACTGGGCTTCCCTTACACCGGA	1500
Db			
QY	1501	GAAGAGGCTTCTGCCAAATACGAGGGGACACTGGAGGCTTAGCAGTGAGGACAG	1560
Db			
QY	1501	GAAGAGGCTTCTGCCAAATACGAGGGGACACTGGAGGCTTAGCAGTGAGGACAG	1560
Db			
QY	1561	CCTGATGACGAGCTTCTGCCAGGCTTAAAGCTTGGAGCTTCCCTTCCCTTAAAGGACAG	1620
Db			
QY	1561	CCTGATGACGAGCTTCTGCCAGGCTTAAAGCTTGGAGCTTCCCTTCCCTTAAAGGACAG	1620
Db			
QY	1621	GGGTGCTGGAGGAGTGGCTGCTCCACCTTCCACCGGCTGCTGGGGGCTTCTGCCCTG	1680
Db			
QY	1621	GGGTGCTGGAGGAGTGGCTGCTCCACCTTCCACCGGCTGCTGGGGGCTTCTGCCCTG	1680
Db			
QY	1681	TGATGCTTCCGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	1740
Db			
QY	1681	TGATGCTTCCGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	1740
Db			
QY	1741	GGGCTGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1800
Db			
QY	1741	GGGCTGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1800
Db			
QY	1801	ATCCCTGTTTATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1860
Db			
QY	1801	ATCCCTGTTTATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1860
Db			
QY	1861	TGCGGAGGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1920
Db			
QY	1861	TGCGGAGGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1920
Db			
QY	1921	CGACTTGGCCAAATACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1980
Db			
QY	1921	CGACTTGGCCAAATACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1980
Db			
QY	1981	CAGTGGTCCAGCTCCCGCTCTGTTAGGCGGCTGGGCTGCGGCTGGGCTGGGCTGGG	2040
Db			
QY	1981	CAGTGGTCCAGCTCCCGCTCTGTTAGGCGGCTGGGCTGCGGCTGGGCTGGGCTGGG	2040
Db			
QY	2041	TTCTGTTGCTGCCAAAGTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2100
Db			
QY	2041	TTCTGTTGCTGCCAAAGTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2100
Db			
QY	2101	GCTGACAGCTGGGGGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2160
Db			
QY	2101	GCTGACAGCTGGGGGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2160
Db			
QY	2161	ACTGGAGGCTTCAAGGGGTTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2220
Db			
QY	2161	ACTGGAGGCTTCAAGGGGTTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2220
Db			
QY	2221	ATGCACTGGAATGCGGGGACTCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2280
Db			

QY 1249 GATGGGAGCCTGGGGCTGTTCTGTCAGTGCAGTCTCCCGGCTTCTCTCTGTCAT 1308
DB 2811 GATGGGAGCCTGGGGCTGTTCTGTCAGTGCAGTCTCCCGGCTTCTCTCTGTCAT 2870
QY 1309 GGACCGGCTGGTGCAGGATTCGGCACTCGACAGTCTATTGGGCGAGTGGGAGCTTT 1368
DB 2871 GGACCGGCTGGTGCAGGATTCGGCACTCGACAGTCTATTGGGCGAGTGGGAGCTTT 2930
QY 1369 CCCTGTGGCTGCCGGTGCACATGCTGTCACAGTGTGGCGCTGGTGCACAGTCTCAGC 1428
DB 2931 CCCTGTGGCTGCCGGTGCACATGCTGTCACAGTGTGGCGCTGGTGCACAGTCTCAGC 2990
QY 1429 CGCCCTACCGGGTTCACCTTCTACGCCCTGCAGATTCCTGCCCTACACACTGGCTCCCT 1488
DB 2991 CGCCCTACCGGGTTCACCTTCTACGCCCTGCAGATTCCTGCCCTACACACTGGCTCCCT 3050
QY 1489 CTACACCGGGAGACAGGCTGTTCTGCGCCAAATACCGAGGGGACACTGGAGTGTCTAG 1548
DB 3051 CTACACCGGGAGACAGGCTGTTCTGCGCCAAATACCGAGGGGACACTGGAGTGTCTAG 3110
QY 1549 CAGTAGGACAGCTGATGACCACTTCTGTCAGGCCCTTAAGCCTGGAGTCCCTTCCC 1608
DB 3111 CAGTAGGACAGCTGATGACCACTTCTGTCAGGCCCTTAAGCCTGGAGTCCCTTCCC 3170
QY 1609 TAATGACACGTGGGTGCTGGAGGACAGTGGCTGCTCCACCTCCACCGCGCTCTGGGG 1668
DB 3171 TAATGACACGTGGGTGCTGGAGGACAGTGGCTGCTCCACCTCCACCGCGCTCTGGGG 3230
QY 1669 GGCTCTGCTGTGATGCTCTCGTAGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1728
DB 3231 GGCTCTGCTGTGATGCTCTCGTAGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3290
QY 1729 GGTTCGGCGCGGGGCTGCTGCTGGACCTCGCCATCTGTCAGTGTGGTGGTGGTGGTGG 1788
DB 3291 GGTTCGGCGCGGGGCTGCTGCTGGACCTCGCCATCTGTCAGTGTGGTGGTGGTGGTGG 3350
QY 1789 CCAGGTGGCCCATCCCTGTTTATGGGTCCATGTCAGCTTCAGCCAGTCTGTCACTGC 1848
DB 3351 CCAGGTGGCCCATCCCTGTTTATGGGTCCATGTCAGCTTCAGCCAGTCTGTCACTGC 3410
QY 1849 CTATATGCTGTGTCGGCAGGCTGGTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 1908
DB 3411 CTATATGCTGTGTCGGCAGGCTGGTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 3470
QY 1909 ATTTGACAGAGCGACTTGGCCAAATACCTACAGCTAGAAAATCTCCAGCACATGGGGTG 1968
DB 3471 ATTTGACAGAGCGACTTGGCCAAATACCTACAGCTAGAAAATCTCCAGCACATGGGGTG 3530
QY 1969 GAGGGCTGCCTCACTGGGTCCAGCTCCCGCTCTGTTAGCCCATGGGGCTCCCGGG 2028
DB 3531 GAGGGCTGCCTCACTGGGTCCAGCTCCCGCTCTGTTAGCCCATGGGGCTCCCGGG 3590
QY 2029 CTGGCCGCGAGTTCTGTTGCTGCCAAGTATGTCGCTCTGTCGTCGTCGTCGTCGTCG 2088
DB 3591 CTGGCCGCGAGTTCTGTTGCTGCCAAGTATGTCGCTCTGTCGTCGTCGTCGTCGTCG 3650
QY 2089 CTGAGTGTGCTAGCTGCAGCTGGGGCTGGGGCTGCTCTCTCTCTCTCTCTCTCTCTCT 2148
DB 3651 CTGAGTGTGCTAGCTGCAGCTGGGGCTGGGGCTGCTCTCTCTCTCTCTCTCTCTCTCT 3710
QY 2149 TAGGGCTGCTGACTGGAGGCTTCCAGGGGGTTCAGTCTGAGTCTATACAGGAGGCT 2208
DB 3711 TAGGGCTGCTGACTGGAGGCTTCCAGGGGGTTCAGTCTGAGTCTATACAGGAGGCT 3770
QY 2209 CAGAAGGCTCCATGACTGGAAATGGGGGACTCTGACGGTGGATTAACAGGCTCAGGG 2268
DB 3771 CAGAAGGCTCCATGACTGGAAATGGGGGACTCTGACGGTGGATTAACAGGCTCAGGG 3830
QY 2269 TTAACAGCTAGCTCTCTAGTCTGACACACACTAGAGAGGGTGTGGAGCTGAATAA 2328
DB 3831 TTAACAGCTAGCTCTCTAGTCTGACACACACTAGAGAGGGTGTGGAGCTGAATAA 3890

QY 2329 CTAGTCACTCGGTGTTCCCATCTCTAAGCCCTTAACTGTCAGCTTCGTTTAATGTAGCT 2388
DB 3891 CTAGTCACTCGGTGTTCCCATCTCTAAGCCCTTAACTGTCAGCTTCGTTTAATGTAGCT 3950
QY 2389 CTTCGATGGGAGTTCCTAGGATGAACACTCCTCCATGGGATTTGAACATATG - ACTTA 2446
DB 3951 CTTCGATGGGAGTTCCTAGGATGAACACTCCTCCATGGGATTTGAACATATGAAGTTA 4010
QY 2447 TTTGTAGGGAGAGTCTCTAGGGGCAACACACAGAACAGGTCCTCCAGCCACAGC 2506
DB 4011 TTTGTAGGGAGAGTCTCTAGGGGCAACACACAGAACAGGTCCTCCAGCCACAGC 4070
QY 2507 ACTGTCTTTTGTGTCATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTGGTCC 2566
DB 4071 ACTGTCTTTTGTGTCATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTGGTCC 4130
QY 2567 TTTGTGTCCTATCAGAGACACAGGATTTAAATATTTAACTTATTTTAAACAAGT 2626
DB 4131 TTTGTGTCCTATCAGAGACACAGGATTTAAATATTTAACTTATTTTAAACAAGT 4190
QY 2627 AGAAGGAATCCATTGCTAGCTTTCTGCTGTGCTCTAATATTTGGTAGGGTGGGG 2686
DB 4191 AGAAGGAATCCATTGCTAGCTTTCTGCTGTGCTCTAATATTTGGTAGGGTGGGG 4250
QY 2687 ATCCCAACAATCAGGTCCTCGATAGTGTGCTATTGGGCTGATCATTCGCCAGAATCT 2746
DB 4251 ATCCCAACAATCAGGTCCTCGATAGTGTGCTATTGGGCTGATCATTCGCCAGAATCT 4310
QY 2747 TCTTCTCTGGGTCCTGGCCCCCAAAATGCTTAACCCAGGACCTTGGAAATTTACTCA 2806
DB 4311 TCTTCTCTGGGTCCTGGCCCCCAAAATGCTTAACCCAGGACCTTGGAAATTTACTCA 4370
QY 2807 TCCCAAAATGATAATTCCAAATGCTTTACCAAGGTTAGGCTGTGAAGGAGTACAGG 2866
DB 4371 TCCCAAAATGATAATTCCAAATGCTTTACCAAGGTTAGGCTGTGAAGGAGTACAGG 4430
QY 2867 GTGGGCTTCAAGGCTTCCCTCAACCGCTTCCCTAACCACCTCTCTCTCTGGCCAGCTTGGT 2926
DB 4431 GTGGGCTTCAAGGCTTCCCTCAACCGCTTCCCTAACCACCTCTCTCTCTGGCCAGCTTGGT 4490
QY 2927 TCCCCCACTTCCACTCCCTCTACTCTCTAGGACTGGGCTGATGAAGGACCTGCCCA 2986
DB 4491 TCCCCCACTTCCACTCCCTCTACTCTCTAGGACTGGGCTGATGAAGGACCTGCCCA 4550
QY 2987 AAATTTCCCTTACCCCAACTTCCCTTACCCCAACTTCCCCAGGCTTCCCAACCC 3046
DB 4551 AAATTTCCCTTACCCCAACTTCCCTTACCCCAACTTCCCCAGGCTTCCCAACCC 4610
QY 3047 TGTGTTGAGCTTACTGCAGGACAGAGCAAAAGTGGGTTTCCCAAGCTTTTGTCCATC 3106
DB 4611 TGTGTTGAGCTTACTGCAGGACAGAGCAAAAGTGGGTTTCCCAAGCTTTTGTCCATC 4670
QY 3107 TCAGCCCGCAGAGTATATCTGCTGGGGAATCTCACAGAACTTCAGAGGACCCCCC 3166
DB 4671 TCAGCCCGCAGAGTATATCTGCTGGGGAATCTCACAGAACTTCAGAGGACCCCCC 4730
QY 3167 TGCTGTAGCTAAGGAGTCTTATCTCTAGGGGGGTTTAAAGTGGCTTGAAGTAAATG 3226
DB 4731 TGCTGTAGCTAAGGAGTCTTATCTCTAGGGGGGTTTAAAGTGGCTTGAAGTAAATG 4790
QY 3227 TCGTCTTATTTATTTAGCGGGTGAATATTTTACTGTAGTGAAGCAATCAGAGTATAA 3286
DB 4791 TCGTCTTATTTATTTAGCGGGTGAATATTTTACTGTAGTGAAGCAATCAGAGTATAA 4850
QY 3287 TGTTTATGGTGACAAAATTAAGGCTTCTTATATGTTTAAAAA 3330
DB 4851 TGTTTATGGTGACAAAATTAAGGCTTCTTATATGTTTAAAAA 4894

RESULT 11
US-09-071-710-16
; Sequence 16, Application us/09071710
; Patent No. 6130043


```
QY 2555 GCCTGTTGGTCTCTGTTGCCATCAGACAGCAGCAATTAATAATTTAACTATTT 2614
Db 1381 GCCTGTTGGTCTCTGTTGCCATCAGACAGCAGCAATTAATAATTTAACTATTT 1440
QY 2615 ATTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTTCTGTTGGTGTCTAAATTTGG 2674
Db 1441 ATTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTTCTGTTGGTGTCTAAATTTGG 1500
QY 2675 GTAGGTGGGGATCCCAACAAATCAGTCCCTGAGATAGTGTGTTGGGTGATCA 2734
Db 1501 GTAGGTGGGGATCCCAACAAATCAGTCCCTGAGATAGTGTGTTGGGTGATCA 1560
QY 2735 TTGCCAGATCTTCTCTCTGTTGGGTCTGCCCCCAAAATGCCCTAACCCAGGACCTGG 2794
Db 1561 TTGCCAGATCTTCTCTCTGTTGGGTCTGCCCCCAAAATGCCCTAACCCAGGACCTGG 1620
QY 2795 AAATCTACTATCCCAATATTAATCCAAATGCTGTGTACCAAGTTAGGGTTGAA 2854
Db 1621 AAATCTACTATCCCAATATTAATCCAAATGCTGTGTACCAAGTTAGGGTTGAA 1680
QY 2855 GGAAGTAGAGGTGGGGCTTCAGTCTCAAGGGCTTCCTTAACCAACCCCTCTCTCTG 2914
Db 1681 GGAAGTAGAGGTGGGGCTTCAGTCTCAAGGGCTTCCTTAACCAACCCCTCTCTCTG 1740
QY 2915 GCCAGCTGGTTCCCCCACTTCCACTCCCTCTACTCTCTCTAGGACTGGGCTGATGA 2974
Db 1741 GCCAGCTGGTTCCCCCACTTCCACTCCCTCTACTCTCTCTAGGACTGGGCTGATGA 1800
QY 2975 AGSACTGCCAAAATTTCCCTAGCCCACTTTCCCTACCCCACTTTCCCACTTTCCCA 3034
Db 1801 AGSACTGCCAAAATTTCCCTAGCCCACTTTCCCTACCCCACTTTCCCACTTTCCCA 1860
QY 3035 GCTCCACACCTGTTTGGAGTACTGAGGACGAGACCAAAAGTGGGTTTCCCAAG 3094
Db 1861 GCTCCACACCTGTTTGGAGTACTGAGGACGAGACCAAAAGTGGGTTTCCCAAG 1920
QY 3095 CPTTCTCATCTCAGCCCCCAGATATATCTGTGCGGAATCTCACAGAAACTC 3154
Db 1921 CPTTCTCATCTCAGCCCCCAGATATATCTGTGCGGAATCTCACAGAAACTC 1980
QY 3155 AGGAGCACCCCTGCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTTAAAGTGGC 3214
Db 1981 AGGAGCACCCCTGCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTTAAAGTGGC 2040
QY 3215 TTTGCAATATGCTGCTTATTTATTTAGGGGGTGAATTTTATATGTAAGTGAGCA 3274
Db 2041 TTTGCAATATGCTGCTTATTTATTTAGGGGGTGAATTTTATATGTAAGTGAGCA 2100
QY 3275 ATCAGAGTATATGTTTATGGTGACAAATTAAGGCTTCTTATATGTTTA 3326
Db 2101 ATCAGAGTATATGTTTATGGTGACAAATTAAGGCTTCTTATATGTTTA 2152

RESULT 12
US-09-525-397-16
; Sequence 16, Application US/09525397
; Patent No. 6252047
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GORDON, JULIAN
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: HODGES, STEVEN C.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: KRATOCHVIL, JON D.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; FOR DETECTING DISEASES OF THE PROSTATE

; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/525,397
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,710
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6083-US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2152 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-525-397-16

Query Match 62.7%; Score 2136.4; DB 4; Length 2152;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 1177 GGGGCTGTACAGGGCGTGGCCAGAGCTGAGCCGGGACCGAGCCGGGAGACACTATGA 1236
Db 1 GGGGCTGTACAGGGCGTGGCCAGAGCTGAGCCGGGACCGAGCCGGGAGACACTATGA 60
QY 1237 TGAAGCGCTTCGGATGGGCGCTGTTCTCTGAGTGCGCCATCTCCCTGGTCTT 1296
Db 61 TGAAGCGCTTCGGATGGGCGCTGTTCTCTGAGTGCGCCATCTCCCTGGTCTT 120
QY 1297 CTCCTGTGTCATGAGACCGGCTGGTGACAGGATTCGGCACTCGAGCAGTCTATTTGGCCAG 1356
Db 121 CTCCTGTGTCATGAGACCGGCTGGTGACAGGATTCGGCACTCGAGCAGTCTATTTGGCCAG 180
QY 1357 TGTGGCAGCTTCCCTGTGGCTGGCGGTGCCACATGCTGTCACACAGTGTGGCGGTGT 1416
Db 181 TGTGGCAGCTTCCCTGTGGCTGGCGGTGCCACATGCTGTCACACAGTGTGGCGGTGT 240
QY 1417 GACAGCTTTCAGCGCCCTCAGCGGTTTCACTTCTTACGCTTCTGAGATCTGCCCCATAC 1476
Db 241 GACAGCTTTCAGCGCCCTCAGCGGTTTCACTTCTTACGCTTCTGAGATCTGCCCCATAC 300
QY 1477 ACTGGCCCTCCCTCTTACACCGGGAGAACAGTGTTCCTGCCCAATTAACGAGGGGACAC 1536
Db 301 ACTGGCCCTCCCTCTTACACCGGGAGAACAGTGTTCCTGCCCAATTAACGAGGGGACAC 360
QY 1537 TGGAGGTGCTAGCAGTGAAGGACAGCTGATGACACAGTTCCTTGGCAGGCCCTTAAGCCCTGG 1596
Db 361 TGGAGGTGCTAGCAGTGAAGGACAGCTGATGACACAGTTCCTTGGCAGGCCCTTAAGCCCTGG 420
QY 1597 AGCTCCCTTCCCTTAATGGACAGTGGGTGCTGAGGAGCAGTGGCTTCTCCCACTCCACC 1656
Db 421 AGCTCCCTTCCCTTAATGGACAGTGGGTGCTGAGGAGCAGTGGCTTCTCCCACTCCACC 480
QY 1657 CGCGCTCTCGGGGGCTCTGCTGTGATGTCTCCGTACGTGTGGTGGGTGAGCCAC 1716
```


FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/850,713

FILING DATE: 02-MAY-1997

ATTORNEY/AGENT INFORMATION:

NAME: Becker, Cheryl L.

REGISTRATION NUMBER: 35,441

REFERENCE/DOCKET NUMBER: 6083, US. P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 847/935-1729

TELEFAX: 847/938-2623

TELEX:

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 2143 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-071-710-15

Query Match 62.08; Score 2114.8; DB 3; Length 2143;
Best Local Similarity 99.88; Pred. No. 0;
Matches 2139; Conservative 0; Mismatches 2; Indels 3; Gaps 2;

QY 1185 ACCAGGCGTGGCCAGAGCTGAGCGGGACCGGACCGGAGACACTATGATGAAGGCG 1244

DB 1 ACCAGGCGTGGCCAGAGCTGAGCGGGACCGGAGACCGGAGACACTATGATGAAGGCG 60

QY 1245 TTCGGATGGCAGCGTGGGCGTGTCTCGAGTGCGCCATCTCCCTGGTCTCTCTCTGG 1304

DB 61 TTCGGATGGCAGCGTGGGCGTGTCTCGAGTGCGCCATCTCCCTGGTCTCTCTCTGG 120

QY 1305 TCATGACCGCGTGGTGCACCGATTGGGCACTCGAGGAGTATTTGGCCAGTGGGCGAG 1364

DB 121 TCATGACCGCGTGGTGCACCGATTGGGCACTCGAGGAGTATTTGGCCAGTGGGCGAG 180

QY 1365 CTTTCCCTGTGGTGGCGGTGCCACATGCTGTCACAGTGTGGCGGTGGTGCACAGCTT 1424

DB 181 CTTTCCCTGTGGTGGCGGTGCCACATGCTGTCACAGTGTGGCGGTGGTGCACAGCTT 240

QY 1425 CAGCGCGCTTCACCGGTTCCACTTCTCAGCCCTCGAGATCTCCCTACACACTGGCCT 1484

DB 241 CAGCGCGCTTCACCGGTTCCACTTCTCAGCCCTCGAGATCTCCCTACACACTGGCCT 300

QY 1485 CCCTCTACACCGGAGAGAGGTGTTCTGCCCCAAATACCCAGGGGACACTGGAGGTG 1544

DB 301 CCCTCTACACCGGAGAGAGGTGTTCTGCCCCAAATACCCAGGGGACACTGGAGGTG 360

QY 1545 CTAGCAGTGGGACAGCGTCTGATACCACTTCTGCGAGCCCTTAAGCCCTGGAGCTCCCT 1604

DB 361 CTAGCAGTGGGACAGCGTCTGATGACCACTTCTGCGAGCCCTTAAGCCCTGGAGCTCCCT 420

QY 1605 TCCCTAATGACACAGTGGGTGCTGGAGGAGTGGCCTGTCTCCACCTCCACCGCGCTCT 1664

DB 421 TCCCTAATGACACAGTGGGTGCTGGAGGAGTGGCCTGTCTCCACCTCCACCGCGCTCT 480

QY 1665 GCGGGGCTCTGCGCTGTGATGTCTCCGTACGTGTGGTGGTGGTGGTGGTGGTGGTGG 1724

DB 481 GCGGGGCTCTGCGCTGTGATGTCTCCGTACGTGTGGTGGTGGTGGTGGTGGTGGTGG 540

QY 1725 GGGTGGTTCGGGGCGGGGATCTGCTGGACCTCGCCATCTCGGATGTCCTCTCTGTC 1784

DB 541 GGGTGGTTCGGGGCGGGGATCTGCTGGACCTCGCCATCTCGGATGTCCTCTCTGTC 600

QY 1785 TGTCCAGGTGGCGCCCTCCCTGTTTATGGGCTCCATTGTCAGCTCAGCCAGTCTGTCA 1844

DB 601 TGTCCAGGTGGCGCCCTCCCTGTTTATGGGCTCCATTGTCAGCTCAGCCAGTCTGTCA 660

QY 1845 CTGCGTATATGTTCTGCGCGGCGGCTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 1904

DB 661 CTGCGTATATGTTCTGCGCGGCGGCTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 720

QY 1905 TAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACCTTCCAGCACATTGG 1964

DB 721 TAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACCTTCCAGCACATTGG 780

QY 1965 GGTGAGGGCGTGCCTCACTGGTCCCGAGCTCCCGGCTCCCTGTTAGCCCCATGGGGCTGC 2024

DB 781 GGTGAGGGCGTGCCTCACTGGTCCCGAGCTCCCGGCTCCCTGTTAGCCCCATGGGGCTGC 840

QY 2025 CGGGCTGGCGCCAGTCTGTGCTGCCAAAGTAATGTGCTCTCTCTCTCTCTCTCTCTCT 2084

DB 841 CGGGCTGGCGCCAGTCTGTGCTGCCAAAGTAATGTGCTCTCTCTCTCTCTCTCTCTCT 900

QY 2085 GCTGCTGAGTGCCTGAGTGCACAGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGG 2144

DB 901 GCTGCTGAGTGCCTGAGTGCACAGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGG 960

QY 2145 TCTCTAGGGTGCCTGAGTGCAGGGGCTTCCAAAGGGGGTTCAGTGTGAGCTTATACAGGG 2204

DB 961 TCTCTAGGGTGCCTGAGTGCAGGGGCTTCCAAAGGGGGTTCAGTGTGAGCTTATACAGGG 1020

QY 2205 AGGCCAGAGGGCTCCATGCAGTGCAGGAGTGCAGGCTGAGTGCAGGCTGAGTGCAGGCTC 2264

DB 1021 AGGCCAGAGGGCTCCATGCAGTGCAGGAGTGCAGGCTGAGTGCAGGCTGAGTGCAGGCTC 1079

QY 2265 AGGGTTAACAGCTAGCTTCCCTAGTGTGAGACACACCTAGAGAGGGTTCCTGGAGCTGAA 2324

DB 1080 AGGGTTAACAGCTAGCTTCCCTAGTGTGAGACACACCTAGAGAGGGTTCCTGGAGCTGAA 1139

QY 2325 TAACTCAGTCACTTGGTTCCTTCCATCTTAAGCCCCCTTAACCTGCGAGCTTCGTTTAAATGT 2384

DB 1140 TAACTCAGTCACTTGGTTCCTTCCATCTTAAGCCCCCTTAACCTGCGAGCTTCGTTTAAATGT 1199

QY 2385 AGCTCTGCATGGGAGTTCCTAGGATGAACACCTCTCCATGGGATTTGAACATATGAA 2442

DB 1200 AGCTCTGCATGGGAGTTCCTAGGATGAACACCTCTCCATGGGATTTGAACATATGAA 1259

QY 2443 CTTTATTTGTAGGGGAGAGTCTCTGAGGGGCAACACACAGAACAGGCTCCCTCAGCCCA 2502

DB 1260 GTTATTTGTAGGGGAGAGTCTCTGAGGGGCAACACACAGAACAGGCTCCCTCAGCCCA 1319

QY 2503 CAGCACTGTCTTTTGTGATCCACCCCTCTTACCTTTTATCAGGATGTGGCCTGTG 2562

DB 1320 CAGCACTGTCTTTTGTGATCCACCCCTCTTACCTTTTATCAGGATGTGGCCTGTG 1379

QY 2563 GTCTCTCTGCTCCCATCAGAGACAGGCACTTTAAATATTTAACTTATTTATTAACA 2622

DB 1380 GTCTCTCTGCTCCCATCAGAGACAGGCACTTTAAATATTTAACTTATTTATTAACA 1439

QY 2623 AAGTAGAGGGAATCCATTGCTAGCTTTTCTGTGTGTGTCTAATATTTGGGTAGGGTG 2682

DB 1440 AAGTAGAGGGAATCCATTGCTAGCTTTTCTGTGTGTGTCTAATATTTGGGTAGGGTG 1499

QY 2683 GGGGATCCCAACATCAGGTCCCTGAGATAGCTTGGCTATGCTATTCCTCAG 2742

DB 1500 GGGGATCCCAACATCAGGTCCCTGAGATAGCTTGGCTATGCTATTCCTCAG 1559

QY 2743 ATCTCTCTCTGCTGGGCTGCTGCCCCCAAAATGCTTAACCCAGGACCTTGAATTTCTA 2802

DB 1560 ATCTCTCTCTGCTGGGCTGCTGCCCCCAAAATGCTTAACCCAGGACCTTGAATTTCTA 1619

QY 2803 CTCATCCCAATATGATAATTTCCAAATGCTTTACCCAAAGGTAGGGTGTGAAGGAAGTA 2862

DB 1620 CTCATCCCAATATGATAATTTCCAAATGCTTTACCCAAAGGTAGGGTGTGAAGGAAGTA 1679

QY 2863 GAGGGTGGGGCTTCAGGTCTCAAGGCTTCCCTTAACCCCTCTCTCTCTCTCTCTCTCT 2922

DB 1680 GAGGGTGGGGCTTCAGGTCTCAAGGCTTCCCTTAACCCCTCTCTCTCTCTCTCTCTCT 1739

QY 2923 TGGTTCCTCCCT 2982

DB 1740 TGGTTCCTCCCT 1799

QY 2983 CCAAAATTTCCCTTACCCCACTTCCCTTACCCCACTTCCCTTACCCCACTTCCCACT 3042

[illegible]

```

RESULT 17
US-09-030-607-10
; Sequence 10, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICATE: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle

```

%	STATE:	WA
%	COUNTRY:	USA
%	ZIP:	98104
%	COMPUTER READABLE FORM:	
%	MEDIUM TYPE:	Floppy disk
%	COMPUTER:	IBM PC compatible
%	OPERATING SYSTEM:	PC-DOS/MS-DOS
%	SOFTWARE:	Patentin Release #1.0, Version #1.30
%	CURRENT APPLICATION DATA:	
%	APPLICATION NUMBER:	US/09/030,607
%	FILING DATE:	25-FEB-1998
%	CLASSIFICATION:	
%	ATTORNEY/AGENT INFORMATION:	
%	NAME:	Maki, David J.
%	REGISTRATION NUMBER:	31,392
%	REFERENCE/DOCKET NUMBER:	210121.427C3
%	TELECOMMUNICATION INFORMATION:	
%	TELEPHONE:	(206) 622-4900
%	TELEFAX:	(206) 682-6031
%	INFORMATION FOR SEQ ID NO:	10:
%	SEQUENCE CHARACTERISTICS:	
%	LENGTH:	789 base pairs
%	TYPE:	nucleic acid
%	STRANDEDNESS:	single
%	TOPOLOGY:	linear
%	MOLECULE TYPE:	cDNA
%	US-09-030-607-10	
%	Query Match	19.7%; Score 673.4; DB 4; Length 789;
%	Best Local Similarity	94.2%; Pred. No. 2.9e-124;
%	Matches	745; Conservative 0; Mismatches 40; Indels 6; Gaps 5;
QY	1341	CAGTCTATTTGGCCACTGTGGCAGCTTCCTCGTGCTGCCGTCACCATGCCTGTCCTC
Db	1	CAGTCTATNTGGCCACTGTGGCAGCTTCCTCGTGCTGCCGTCACCATGCCTGTCCTC
QY	1401	ACAGTGTGGCCGTGGTGACAGCTTCAGCGGCCCTCACCGGGTTCACTTCTCAGGCCCTGC
Db	61	ACAGTGTGGCCGTGGTGACAGCTTCAGCGGCCCTCACCGGGTTCACTTCTCAGGCCCTGC
QY	1461	AGATCTCTGGCCCTACACACTGGCCCTCCCTCTACACCGGGAGAAGCAGGTGTTCTCGCCCA
Db	121	AGATCTCTGGCCCTACACACTGGCCCTCCCTCTACACCGGGAGAAGCAGGTGTTCTCGCCCA
QY	1521	AATACCAGGGGACACTGGAGTGCTAGCAGTGAGGACAGCCTGATGACCACTTCCTCTGC
Db	181	AATACCAGGGGACACTGGAGTGCTAGCAGTGAGGACAGCCTGATGACCACTTCCTCTGC
QY	1581	CAGGCCCTAAGCTTGAGCTCCCTTCCTTAATGGACAGTGGTGCTGGAGCAGTGCCC
Db	241	CAGGCCCTAAGCTTGAGCTCCCTTCCTTAATGGACAGTGGTGCTGGAGCAGTGCCC
QY	1641	TGCTCCCACTCCACCCGCTCTCGGGGCCCTCTGCCTGTGATGTCCTCCGTACCTGTGG
Db	301	TGCTCCCACTCCACCCGCTCTCGGGGCCCTCTGCCTGTGATGTCCTCCGTACCTGTGG
QY	1701	TGTTGGGTGAGCCCCAGCCAGCGGTGTTCCGGCCGGGGCATCTGCTGTGACACTCG
Db	361	TGTTGGGTGAGCCCCAGCCAGCGGTGTTCCGGCCGGGGCATCTGCTGTGACACTCG
QY	1761	CCATCTGTGATAGTGTCTTCCTGTCTCCAGTGGCCCACTCCCTGTTTTATGGGCTCCA
Db	421	CCATCTGTGATAGTGTCTTCCTGTCTCCANGTGGCCCACTCCCTGTTTTATGGGCTCCA
QY	1821	TTGTCCAGCTCAGCAGTCTGTCACTGCTATPATNGTGTCTGTGCCAGAGCCCTGGGTCTGG
Db	480	TTGTCCAGCTCAGCAGTCTGTCACTGCTATPATNGTGTCTGTGCCAGAGCCCTGGGTCTGG
QY	1881	TCGCCATTTTACTTTCGTACAGGTAGTATTTTGACAAGAGCAGCTTGCCCAAATACTCAG
Db	540	TC-CCAATTTACTTTTGTCTACAGGTANTATTTTGACAAGAAGAGNTTGGCCAAATACTCAG
QY	1941	CGTAGAAAATTTCCAGCA--CATTTGGGTGGAGGGCCCTGCCTCACTGGGTCCCACTGCC
QY		1998

Db 599 CGTAAATTTCCAGACATTGGGGTGGAGGCTTCCCTCACTGGGT-CCAACTCCC 657
Qy 1999 CGCTCTCTAGCCCATGGGCTGCGGGCTGGCGGCGGAGTTTCTGTGTGCTGCCAAAGT 2058
Db 658 CGCTCTCTTAAACCCCATGGGCTGCGGGCTTGGCGGCGCAATTTCTGTGTGCCAAANT 717
Qy 2059 AATGTGGCTCTGCTGCCACCTGTGCT-GCTGAGGTGCTGAGCTGCACAGCTGGGGC 2117
Db 718 NATGTGGCTCTGCTGCCACCTGTGCTGAGTGCNTACNGCNCANTNGGGGG 777
Qy 2118 TGGGGCTGCC 2128
Db 778 TNGGGNGTCC 788

RESULT 18

US-09-605-785-10
; Sequence 10, Application US/09605785
; Patent No. 6321716

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785

; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 789

; TYPE: DNA
; ORGANISM: Homo sapien
; NAME/KEY: misc_feature
; LOCATION: (1)...(789)
; OTHER INFORMATION: n = A,T,C or G

Query Match 19.78; Score 673.4; DB 4; Length 789;
Best Local Similarity 94.28; Pred. No. 2.9e-124;
Matches 745; Conservative 0; Mismatches 40; Indels 6; Gaps 5;

Qy 1341 CAGTCTATTGGCCAGTGGCAGCTTTCCTGTGGCTGCGGTCGCACATGCTGCC 1400
Db 1 CAGTCTATNNGCCAGTGGCAGCTTTCCTGTGGCTGCGGTCGCACATGCTGCC 60
Qy 1401 ACAGTGTGGCGGTGGTGCAGCTTTCAGCGCCCTTACCGGGTTCACCTTCTAGCCCTGC 1460
Db 61 ACAGTGTGGCGGTGGTGCAGCTTTCAGCGCCCTTACCGGGTTCACCTTCTAGCCCTGC 120
Qy 1461 AGATCTGCCCTACACACTGGCCTCCCTTACACCGGGAGAGGAGTCTTCTGCCCA 1520
Db 121 AGATCTGCCCTACACACTGGCCTCCCTTACACCGGGAGAGGAGTCTTCTGCCCA 180
Qy 1521 AATACCGGGGACACTGGAGGTGCTAGCAGTGGAGGAGCCTGATGACAGCTTCTCTGC 1580

Db 181 AATACCAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCCTGATGACACAGCTTCTCTGC 240
Qy 1581 CAGGCCCTAAGCCCTGGAGCTCCCTTCCCTAATGACACAGTGGGTGCTGGAGGAGTGGCC 1640
Db 241 CAGGCCCTAAGCCCTGGAGCTCCCTTCCCTAATGACACAGTGGGTGCTGGAGGAGTGGCC 300
Qy 1641 TGTCTCCACCTCCACCCGCGCTCTGCGGGGCTCTGCTGTGATGTCTCCGTACGTGTGG 1700
Db 301 TGTCTCCACCTCCACCCGCGCTCTGCGGGGCTCTGCTGTGATGTCTCCGTACGTGTGG 360
Qy 1701 TGTGGGTGAGCCACCGAGGAGGCTTCCGGGCGGGGATCTGCTGGAGCTGC 1760
Db 361 TGTGGGTGAGCCACCGAGGAGGCTTCCGGGCGGGGATCTGCTGGAGCTGC 420
Qy 1761 CCATCTCTGGATAGTGTCTCTGCTGCCAGGTGGCCCATCCCTGTTTATGGGTCCA 1820
Db 421 CCATCTCTGGATAGTGTCTCTGCTGCCAGGTGGCCCATCCCTGTTTATGGGTCCA 479
Qy 1821 TTGTCCAGCTCAGCCAGTGTCTACTGCTATATGTTGCTGCCGAGGCTGGGTCTGG 1880
Db 480 TTGTCCAGCTCAGCCAGTGTCTACTGCTATATGTTGCTGCCGAGGCTGGGTCTGG 539
Qy 1881 TCGCCATTCTTCTGTACACAGCTAGTATTTGACAGAGGAGCTTGGCCAAATACCTCAG 1940
Db 540 TC-CCATTCTTCTGTACACAGTANTATTTGACAGAGAGGANTTGGCCAAATACCTCAG 598
Qy 1941 CGTAAACCTTCCAGCA--CATTGGGGTGGAGGCTGCTCTACTGGTCCCGAGTCCC 1998
Db 599 CGTTAAAAAATTCACGCAACATTTGGGGTGGAGGCTTGGCTCAGTCCGCT-CCAACTCCC 657
Qy 1999 CGCTCTCTTAGCCCCCATGGGGCTGCGGGCTGCGGCGGAGTTCCTGTGCTGCCAAAGT 2058
Db 658 CGCTCTCTTAAACCCCATGGGGCTGCGGGCTGCGGCGGAGTTCCTGTGCTGCCAAANT 717
Qy 2059 AATGTGGCTCTGCTGCCACCTGTGCT-GCTGAGGTGCTGAGTGCACAGCTGGGGC 2117
Db 718 NATGTGGCTCTGCTGCCACCTGTGCTGAGTGCNTACNGCNCANTNGGGGG 777
Qy 2118 TGGGGCTGCC 2128
Db 778 TNGGGNGTCC 788

RESULT 19

US-09-439-313-10
; Sequence 10, Application US/09439313
; Patent No. 6329505

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9

; CURRENT APPLICATION NUMBER: US/09/439,313
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 789

; TYPE: DNA
; ORGANISM: Homo sapien
; NAME/KEY: misc_feature
; LOCATION: (1)...(789)

OTHER INFORMATION: n = A,T,C or G
US-09-439-313-10

Query Match 19.7%; Score 673.4; DB 4; Length 789;
Best Local Similarity 94.2%; Pred. No. 2.9e-124;
Matches 745; Conservative 0; Mismatches 40; Indels 6; Gaps 5;

QY 1341 CAGTCTATTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGCCGTCGCCACATGCTGTCCC 1400
DB 1 CAGTCTATNTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGCCGTCGCCACATGCTGTCCC 60

QY 1401 ACAGTGTGGCCGTGTGACAGCTTCAGCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 1460
DB 61 ACAGTGTGGCCGTGTGACAGCTTCAGCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 120

QY 1461 AGATCTGCTTACACACTGGCTTCCCTTACCACCGGGAGAGAGAGTGTTCCTGTGCCA 1520
DB 121 AGATCTGCTTACACACTGGCTTCCCTTACCACCGGGAGAGAGAGTGTTCCTGTGCCA 180

QY 1521 AATACCGAGGGGACACTGGAGTGTAGCAGTGAAGACAGCTGATGACAGCTTCCCTGC 1580
DB 181 AATACCGAGGGGACACTGGAGTGTAGCAGTGAAGACAGCTGATGACAGCTTCCCTGC 240

QY 1581 CAGGCCCTAAGCCTGGAGTGTCCCTTAATGGACAGTGGGTGCTGGAGGCAAGTGGCC 1640
DB 241 CAGGCCCTAAGCCTGGAGTGTCCCTTAATGGACAGTGGGTGCTGGAGGCAAGTGGCC 300

QY 1641 TGCTCCCACTTCCACCGCCGCTCTCGGGGGCTCTGCGGGGCTCTGCGGTGATGTCTCCGTAGTGTGG 1700
DB 301 TGCTCCCACTTCCACCGCCGCTCTCGGGGGCTCTGCGGGGCTCTGCGGTGATGTCTCCGTAGTGTGG 360

QY 1701 TGGTGGGTGAGCCCAACCGAGCCAGGTGTTCGGGGCTGGGGCTATGCTGCGAGCTTCG 1760
DB 361 TGGTGGGTGAGCCCAACCGAGCCAGGTGTTCGGGGCTGGGGCTATGCTGCGAGCTTCG 420

QY 1761 CCATCTCGATAGTGTCTTCTGCTCCAGGTGGCCCCCATTCCTGTATGGGTCCA 1820
DB 421 CCATCTCGATAGTGTCTTCTGCTCCAGGTGGCCCCCATTCCTGTATGGGTCCA 479

QY 1821 TTGTCACAGTGTGTCTACAGTGTATTTGACAGAGGAGTGTGGCCAAATACCTCAG 1940
DB 540 TC-CCATTTACTTTGCTACACAGTGTATTTGACAGAGGAGTGTGGCCAAATACCTCAG 598

QY 1941 CGTAGAAAACCTCCAGCA--CATTTGGGTGGAGGGCTGCCCTACCTGGGTCCAGCTCC 1998
DB 599 CGTTAAAAATTCACACACATTTGGGGGTGGAGGGCTGCCCTACCTGGGT-CCAACTCC 657

QY 2059 AATGTGGCTCTCTGTGCGCCACCTGTGCT-GCTGAGGTGGGTAGCTGCACAGCTGGGGG 2117
DB 718 NATGTGGCTCTCTGTGCGCCACCTGTGCTGGCTGAAGTGCNTACNGCNCANTNGGGGG 777

QY 2118 TGGGCGCTCC 2128
DB 778 TNGGNGTTC 788

RESULT 20

US-09-352-616A-10
Sequence 10, Application US/09352616A
Patent No. 6395278
GENERAL INFORMATION:
APPLICANT: Dillon, Davin C.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang, Yuqi
APPLICANT: Xu, Jiangchun

APPLICANT: Mitcham, Jennifer Lynn
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.427C8
CURRENT APPLICATION NUMBER: US/09/352.616A
CURRENT FILING DATE: 1999-07-13
NUMBER OF SEQ ID NOS: 472
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 10
LENGTH: 789
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(789)
OTHER INFORMATION: n = A,T,C or G
US-09-352-616A-10

Query Match 19.7%; Score 673.4; DB 4; Length 789;
Best Local Similarity 94.2%; Pred. No. 2.9e-124;
Matches 745; Conservative 0; Mismatches 40; Indels 6; Gaps 5;

QY 1341 CAGTCTATTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGCCGTCGCCACATGCTGTCCC 1400
DB 1 CAGTCTATNTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGCCGTCGCCACATGCTGTCCC 60

QY 1401 ACAGTGTGGCCGTGTGACAGCTTCAGCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 1460
DB 61 ACAGTGTGGCCGTGTGACAGCTTCAGCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 120

QY 1461 AGATCTGCTTACACACTGGCTTCCCTTACCACCGGGAGAGAGTGTTCCTGTGCCA 1520
DB 121 AGATCTGCTTACACACTGGCTTCCCTTACCACCGGGAGAGAGTGTTCCTGTGCCA 180

QY 1521 AATACCGAGGGGACACTGGAGTGTAGCAGTGAAGACAGCTGATGACAGCTTCCCTGC 1580
DB 181 AATACCGAGGGGACACTGGAGTGTAGCAGTGAAGACAGCTGATGACAGCTTCCCTGC 240

QY 1581 CAGGCCCTAAGCCTGGAGTGTCCCTTAATGGACAGTGGGTGCTGGAGGCAAGTGGCC 1640
DB 241 CAGGCCCTAAGCCTGGAGTGTCCCTTAATGGACAGTGGGTGCTGGAGGCAAGTGGCC 300

QY 1641 TGCTCCCACTTCCACCGCCGCTCTCGGGGGCTCTGCGGGGCTCTGCGGTGATGTCTCCGTAGTGTGG 1700
DB 301 TGCTCCCACTTCCACCGCCGCTCTCGGGGGCTCTGCGGGGCTCTGCGGTGATGTCTCCGTAGTGTGG 360

QY 1701 TGGTGGGTGAGCCCAACCGAGCCAGGTGTTCGGGGCTGGGGCTATGCTGCGAGCTTCG 1760
DB 361 TGGTGGGTGAGCCCAACCGAGCCAGGTGTTCGGGGCTGGGGCTATGCTGCGAGCTTCG 420

QY 1761 CCATCTCGATAGTGTCTTCTGCTCCAGGTGGCCCCCATTCCTGTATGGGTCCA 1820
DB 421 CCATCTCGATAGTGTCTTCTGCTCCAGGTGGCCCCCATTCCTGTATGGGTCCA 479

QY 1821 TTGTCACAGTGTGTCTACAGTGTATTTGACAGAGGAGTGTGGCCAAATACCTCAG 1880
DB 480 TTGTCACAGTGTGTCTACAGTGTATTTGACAGAGGAGTGTGGCCAAATACCTCAG 539

QY 1881 TCCCAATTTACTTTGCTACACAGTGTATTTGACAGAGGAGTGTGGCCAAATACCTCAG 1940
DB 540 TC-CCATTTACTTTGCTACACAGTGTATTTGACAGAGGAGTGTGGCCAAATACCTCAG 598

QY 1941 CGTAGAAAACCTCCAGCA--CATTTGGGTGGAGGGCTGCCCTACCTGGGTCCAGCTCC 1998
DB 599 CGTTAAAAATTCACACACATTTGGGGGTGGAGGGCTGCCCTACCTGGGT-CCAACTCC 657

QY 2059 AATGTGGCTCTCTGTGCGCCACCTGTGCT-GCTGAGGTGGGTAGCTGCACAGCTGGGGG 2117
DB 718 NATGTGGCTCTCTGTGCGCCACCTGTGCTGGCTGAAGTGCNTACNGCNCANTNGGGGG 777

[illegible]

Db	675	TN	GGGTGCGGGGGTTGCCCCCNAGTTTTTGTGTGTGCCAAAGTAAAGTGGTTTTTTT	616
Qy	2072	GCT	GCCACACCTGTGCTGCTGAGGTGCGTAGCTGCACAGCTGGGGCGTGGGGCGTCCCTCT	2131
Db	615	GCT	GCCACACCTGTGCTGTGAGGTGCGTAGTGCACAGCTGGGGNTGGGGCGTCCCTTTT	556
Qy	2132	CCT	CTCTCTCCCACTCTCTAGGGCTGCGTAGCTGGAGGCCCTTCCAAGGGGTTTCAGTCTG	2191
Db	555	CT	TTTTTCCCACTTTNTAGGGCTGCGTAGCTGGAGNCCCTTCCAAGGGGTTTCAGTCTG	496
Qy	2192	GACT	TATACAGGAGGCCAGAAGGGTCCCATGTCACATGGAATGCGGGGACTCTCGACGGTGG	2251
Db	495	GACT	TATACAGGAGGCCAGAGGGTTCATGCATCGGAATGCGGGGACTCTCGACGGTGG	436
Qy	2252	ATT	ACCCAGGCTCAGGGTTAAACAGCTAGCGTCTCTAGTTGAGACACACTTAGAAGAGGGTT	2311
Db	435	ATT	ACCCAGGCTCAGGGTTAAACAGCTAGCGTCTCTAGTTGAGACACACTTAGAAGAGGGTT	376
Qy	2312	TTT	GGGAGCTGAATAAATCAGTCACTGGTGTTCCTCATCTTAAGCCCTTAACTTCGACG	2371
Db	375	TTT	GGGAGCTGAATAAATCAGTCACTGGTGTTCCTCATCTNTAAGCCCTTAACTTCGACG	316
Qy	2372	CTT	CGTTTTAATGTAGCTCTTGCACTGGAGGTTCTAGGATGAACACTCTCTCCATGGGATT	2431
Db	315	CTT	CGTTTTAATGTAGCTCTTGCACTGGAGGTTCTAGGATGAACACTCTCTCCATGGGATT	256
Qy	2432	TGA	ACATATG--ACTTATTGTAGGGAAGAGTCTTGAGGGGCAACACACAAACACGAG	2489
Db	255	TGA	ACATATGAAGTTATTGTAGGGAAGAGTCTTGAGGGGCAACACACAAACACGAG	196
Qy	2490	TCC	CCCTCAGCCCAAGCACTGTCTTTTGCTGATCCACCCCTCTTACCTTTATCAGG	2549
Db	195	TCC	CCCTCAGCCCAAGCACTGTCTTTTGCTGATCCACCCCTCTTACCTTTATCAGG	136
Qy	2550	ATG	TGGCTGTTGGTCTCTGTGTGGCATCACAGACACAGGCAATTAATAATTTAACT	2609
Db	135	ATG	TGGCTGTTGGTCTCTGTGTGGCATCACAGACACAGGCAATTAATAATTTAACT	76
Qy	2610	TAT	TTTATTAAACAAAGTAGAAGGAATCCATCTAGCTTTCTGTGTGTGGTCTCTAATA	2669
Db	75	TAT	TTTATTAAACAAAGTAGAAGGAATCCATCTAGCTTTCTGTGTGTGGTCTCTAATA	16
Qy	2670	TTT	GGGTAGGGTGGG	2684
Db	15	TTT	GGGTAGGGTGGG	1

```

RESULT 27
US-09-232-149A-11/c
; Sequence 11, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232.149A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 772
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(772)
; OTHER INFORMATION: n = A,T,C or G
US-09-232-149A-11

```


Qy	2712	ATAGCTGTGTCATTTGGGCTGATCATTTGCCAGAAATCTTCTTCTCCTGGGCTCTGGCCCCCA	2771
Db	181	ATAGCTGTGTCATTTGGGCTGATCATTTGCCAGAAATCTTCTTCTCCTGGGCTCTGGCCCCCA	240
Qy	2772	AAATGCCTTAACCCAGGACCTTGGAAATCTTACTCATCCCAATGATAA	2819
Db	241	AAATGCCTTAACCCAGGACCTTGGAAATCTTACTCATCCCAATGATAA	288

RESULT 34

US-09-071-710-11
Sequence 11, Application US/09071710
Patent No. 6130043
GENERAL INFORMATION:
APPLICANT: BILLING-MEDEL, PATRICIA
APPLICANT: COHEN, MAURICE
APPLICANT: COLPITTS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: GORDON, JULIAN
APPLICANT: GRANADOS, EDWARD N.
APPLICANT: HODGES, STEVEN C.
APPLICANT: KASS, MICHAEL R.
APPLICANT: KRATOCHVIL, JON D.
APPLICANT: ROBERTS-RAPP, LISA
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STROUPE, STEPHEN D.
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
TITLE OF INVENTION: FOR DETECTING DISEASES OF THE PROSTATE
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:


```

; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,710
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/850,713
; FILING DATE: 02-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6083.US.PI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-071-710-1

Query Match 7.5%; Score 256.4; DB 3; Length 258;
Best Local Similarity 99.6%; Pred. No. 4e-42; Mismatches 1; Indels 0; Gaps 0;
Matches 257; Conservative 0;

QY 1177 GGGGCTGTACCAAGGCGTGCACAGACTGACGCGGCGACCGAGGCCCGGAGACACTATGA 1236
DB 1 GGGGCTGTACCAAGGCGTGCACAGACTGACGCGGCGACCGAGGCCCGGAGACACTATGA 60

QY 1237 TGAAGCGGTTCCGATGGGAGCGCTGGGGCTGTTTCCTGTCAGTGGCCCATCTCCCTGGTCTT 1296
DB 61 TGAAGCGGTTCCGATGGGAGCGCTGGGGCTGTTTCCTGTCAGTGGCCCATCTCCCTGGTCTT 120

QY 1297 CTCCTCTGTCATGGACCGCGTGTGTCAGGCGATTTCGGCAGCTCGAGCAGTCTATTTGGCCAG 1356
DB 121 CTCCTCTGTCATGGACCGCGTGTGTCAGGCGATTTCGGCAGCTCGAGCAGTCTATTTGGCCAG 180

QY 1357 TGTGGCAGCTTTCCCTGTGGCTGCCGTCGCCACATGCCTGTGCCACAGTGTGCCCGTGGT 1416
DB 181 TGTGGCAGCTTTCCCTGTGGCTGCCGTCGCCACATGCCTGTGCCACAGTGTGCCCGTGGT 240

QY 1417 GACAGCTTCAGCGCCCT 1434
DB 241 GACAGCTTCAGCGCCCT 258

RESULT 39
US-09-525-397-1
; Sequence 1, Application US/09525397
; Patent No. 6252047
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GORDON, JULIAN
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: HODGES, STEVEN C.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: KRATOCHVIL, JON D.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL

```

TITLE OF INVENTION: FOR DETECTING DISEASES OF THE PROSTATE

NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/525,397
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/071,710
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6083.US.P1
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 258 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-525-397-1

Query Match 7.5%; Score 256.4; DB 4; Length 258;

Best Local Similarity 99.6%; Pred. No. 4e-42;

Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1177 GGGCGCTGACAGGCGCTGCCAGAGTGGCCAGGCGGACACCGAGGCGGAGACACTATGA 1236
Db 1 GGGCGCTGACAGGCGCTGCCAGAGTGGCCAGGCGGACACCGAGGCGGAGACACTATGA 60

QY 1237 TGAAGCGCTTCGGATGGGCGAGCGTGGGCTGTTCTTCAGTGCACATCTCCCTGGCTT 1296
Db 61 TGAAGCGCTTCGGATGGGCGAGCGTGGGCTGTTCTTCAGTGCACATCTCCCTGGCTT 120

QY 1297 CTCTCTGTGATGACCGGCTGGTGACGAGTTCGGCACTCGACAGTCTATTTGGCCAG 1356
Db 121 CTCTCTGTGATGACCGGCTGGTGACGAGTTCGGCACTCGACAGTCTATTTGGCCAG 180

QY 1357 TGTGACAGCTTCCCTGTGGCTGGCGGTCGACATGCTGTCACAGTGGCCGCTGGT 1416
Db 181 TGTGACAGCTTCCCTGTGGCTGGCGGTCGACATGCTGTCACAGTGGCCGCTGGT 240

QY 1417 GACAGCTTCAGCGCGCT 1434

Db 241 GACAGCTTCAGCGCGCT 258

RESULT 40

US-09-071-710-3

Sequence 3, Application US/09071710

Patent No. 6130043

GENERAL INFORMATION:

APPLICANT: BILLING-MEDEL, PATRICIA

APPLICANT: COHEN, MAURICE

APPLICANT: COLPITTS, TRACEY L.

APPLICANT: FRIEDMAN, PAULA N.

APPLICANT: GORDON, JULIAN

APPLICANT: GRANADOS, EDWARD N.

APPLICANT: HODGES, STEVEN C.
APPLICANT: KLASS, MICHAEL R.
APPLICANT: KRATOCHVIL, JON D.
APPLICANT: ROBERTS-RAPP, LISA
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STROUPE, STEPHEN D.
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
FOR DETECTING DISEASES OF THE PROSTATE
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,710
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/850,713
FILING DATE: 02-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6083.US.P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 255 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: base_polymorphism

LOCATION: 215

OTHER INFORMATION: /note= " N' represents an A or G or

T or C polymorphism at this position"

US-09-071-710-3

Query Match

Best Local Similarity 7.4%; Score 254; DB 3; Length 255;

Matches 254; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1417 GACAGCTTCAGCGCGCTCACCGGTTACCTTCTCAGCCCTGCAGATCTTCCCTACAC 1476
Db 1 GACAGCTTCAGCGCGCTCACCGGTTACCTTCTCAGCCCTGCAGATCTTCCCTACAC 60

QY 1477 ACTGGCTCCCTCTACACCGGAGAGAGAGTGTCTTCTGCCCAATACCGAGGGACAC 1536
Db 61 ACTGGCTCCCTCTACACCGGAGAGAGAGTGTCTTCTGCCCAATACCGAGGGACAC 120

QY 1537 TGGAGGTGCTAGCAGTGGAGACAGCTGATGACAGCTTCTCCAGGCGCTTAAGCTGG 1596
Db 121 TGGAGGTGCTAGCAGTGGAGACAGCTGATGACAGCTTCTCCAGGCGCTTAAGCTGG 180

QY 1597 AGCTCCCTTCCCTTAATGGACACGCTGGGTGCTGGAGGAGTGGCTCTCCACCTCCACC 1656
Db 181 AGCTCCCTTCCCTTAATGGACACGCTGGGTGCTGGAGGAGTGGCTCTCCACCTCCACC 240

QY 1657 CGCGCTCTGGGGGC 1671

Db 241 CGCGCTCTGGGGGC 255

RESULT 41
US-09-525-397-3
: Sequence 3, Application US/09525397.
: Patent No. 6252047
: GENERAL INFORMATION:
: APPLICANT: BILLING-MEDEL, PATRICIA
: APPLICANT: COHEN, MAURICE
: APPLICANT: COLPITTS, TRACEY L.
: APPLICANT: FRIEDMAN, PAULA N.
: APPLICANT: GORDON, JULIAN
: APPLICANT: GRANADOS, EDWARD N.
: APPLICANT: HODGES, STEVEN C.
: APPLICANT: KLASS, MICHAEL R.
: APPLICANT: KRATOCHVIL, JON D.
: APPLICANT: ROBERTS-RAPP, LISA
: APPLICANT: RUSSELL, JOHN C.
: APPLICANT: STROUPE, STEPHEN D.
: TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
: FOR DETECTING DISEASES OF THE PROSTATE
: NUMBER OF SEQUENCES: 41
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Abbott Laboratories
: STREET: 100 Abbott Park Road
: CITY: Abbott Park
: STATE: IL
: COUNTRY: USA
: ZIP: 60064-3500
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/525,397
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/071,710
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Becker, Cheryl L.
: REGISTRATION NUMBER: 35,441
: REFERENCE/DOCKET NUMBER: 6083.US.P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 847/935-1729
: TELEFAX: 847/938-2623
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 255 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: base_polymorphism
: LOCATION: 215
: OTHER INFORMATION: /note= " N' represents an A or G or
: OTHER INFORMATION: T or C polymorphism at this position "
US-09-525-397-3
Query Match 7.4%; Score 254; DB 4; Length 255;
Best Local Similarity 99.6%; Pred. No. 1.2e-41;
Matches 254; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1417 GACAGCTTCAGCGCGCCCTCACCGGTTTCACCTTCTCAGCCCTGCAGATCTCTGCCCTACAC 1476
Db 1 GACAGCTTCAGCGCGCCCTCACCGGTTTCACCTTCTCAGCCCTGCAGATCTCTGCCCTACAC 60
QY 1477 ACTGGCTTCCTCTACACCGGAGACAGAGTGTTCCTGCCCAAAATACCGAGGGGACAC 1536
Db 61 ACTGGCTTCCTCTACACCGGAGAGAGAGTGTTCCTGCCCAAAATACCGAGGGGACAC 120

QY 1537 TGGAGGTGCTAGCAGTGGAGACAGCTGATGACACAGCTTCCTGCCAGGCCCTTAAGCCTGG 1596
Db 121 TGGAGGTGCTAGCAGTGGAGACAGCTGATGACACAGCTTCCTGCCAGGCCCTTAAGCCTGG 180
QY 1597 AGCTCCCTTCCCTAATGACACGTGGTCTGAGGACAGTGGCCTGCTCCACACCTCCACC 1656
Db 181 AGCTCCCTTCCCTAATGACACGTGGTCTGAGGACAGTGGTCTGAGGACAGTGGCCTGCTCCACACCTCCACC 240
QY 1657 CGCGCTCTCGGGGC 1671
Db 241 CGCGCTCTCGGGGC 255
RESULT 42
US-09-071-710-4
: Sequence 4, Application US/09071710
: Patent No. 6130043
: GENERAL INFORMATION:
: APPLICANT: BILLING-MEDEL, PATRICIA
: APPLICANT: COHEN, MAURICE
: APPLICANT: COLPITTS, TRACEY L.
: APPLICANT: FRIEDMAN, PAULA N.
: APPLICANT: GORDON, JULIAN
: APPLICANT: GRANADOS, EDWARD N.
: APPLICANT: HODGES, STEVEN C.
: APPLICANT: KLASS, MICHAEL R.
: APPLICANT: KRATOCHVIL, JON D.
: APPLICANT: ROBERTS-RAPP, LISA
: APPLICANT: RUSSELL, JOHN C.
: APPLICANT: STROUPE, STEPHEN D.
: TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
: FOR DETECTING DISEASES OF THE PROSTATE
: NUMBER OF SEQUENCES: 41
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Abbott Laboratories
: STREET: 100 Abbott Park Road
: CITY: Abbott Park
: STATE: IL
: COUNTRY: USA
: ZIP: 60064-3500
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/071,710
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/850,713
: FILING DATE: 02-MAY-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Becker, Cheryl L.
: REGISTRATION NUMBER: 35,441
: REFERENCE/DOCKET NUMBER: 6083.US.P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 847/935-1729
: TELEFAX: 847/938-2623
: TELEX:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 247 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-071-710-4
Query Match 7.2%; Score 247; DB 3; Length 247;
Best Local Similarity 100.0%; Pred. No. 2.8e-40;
Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1598 GCTCCCTTCCCTAATGACACAGTGGTCTGAGGACAGTGGCCTGCTCCACACCTCCACC 1657

D _b	1	GCTCCCTTCCTTAATGGACAGTGGGTGCTGGAGGCAGTGCGCTGCTCCACACTCCACC	60
Q _y	1658	GCGCCTGCGGGGCCCTCTGCCTGTGATGTTCCGTACGTGTGTGTGGGTGAGCCCAACC	1717
D _b	61	GCCTCTGCGGGGCCCTTGCTGTGATGTTCCGTACGTGTGGTGGGTGAGCCCAACC	120
Q _y	1718	GAGGCCAAGGTGTTTCCGGSCGGGGCATCTGCTCGACCTCGCCACTCCTTGGATAGTGCC	1777
D _b	121	GAGGCCAAGGTGTTTCCGGSCGGGGCATCTGCTCGACCTCGCCACTCCTTGGATAGTGCC	180
Q _y	1778	TTCCTGCTGCCAGGTGGCCCCATCCCTGTTTTATGGGCTCCATTGTCCAGCTCAGCCAG	1837
D _b	181	TTCTGCTGTCCCAGGTGGCCCCATCCCTGTTTTATGGGCTCCATTGTCCAGCTCAGCCAG	240
Q _y	1838	TCGTCA	1844
D _b	241	TCGTCA	247

RESULT 43
US-09-525-397-4
; Sequence 4, Application US/09525397
; Patent No. 6252047
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GORDON, JULIAN
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: HODGES, STEVEN C.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: KRATOCHVIL, JON D.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; FOR DETECTING DISEASES OF THE PROSTATE
; NUMBER OF SEQUENCES: 41

```

US-09-525-397-4

Query Match          7.2%; Score 247; DB 4; Length 247;
Best Local Similarity 100.0%; Pred. No. 2.8e-40;
Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1598  GCTCCCTTCCCTAATGGACACAGTGGGTGCTGGAGGCAGTGGCCTGCTCCACCTCCACCC 1557
          |||||
Db       1    GCTCCCTTCCCTAATGGACACAGTGGGTGCTGGAGGCAGTGGCCTGCTCCACCTCCACCC 60

QY      1658  GCGCTCTCGCGGGCCCTGCTGCTGATGTCTCCGTACGTGTGGTGGGTGAGCCACCC 1717
          |||||
Db       61   GCGCTCTCGCGGGCCCTGCTGCTGATGTCTCCGTACGTGTGGTGGGTGAGCCACCC 120

QY      1718  GAGGCCAGGGTGTTCGCCGGCGGGGCATCTGCCTGGACTCGCCATCTCGGATATGTC 1777
          |||||
Db       121  GAGGCCAGGGTGTTCGCCGGCGGGGCATCTGCCTGGACTCGCCATCTCGGATATGTC 180

QY      1778  TTCCTCTGCTGCCAGGTGCCCCATCCCTGTTTATGGCTGCCATTCGACGCTCAGCCAG 1837
          |||||
Db       181  TTCCTCTGCTGCCAGGTGCCCCATCCCTGTTTATGGCTGCCATTCGACGCTCAGCCAG 240

QY      1838  TCTGTCA 1844
          |||||
Db       241  TCTGTCA 247

RESULT 44
US-09-602-877A-93/c
; Sequence 93, Application US/09602877A
; Patent No. 6432707
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun G.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.446C5
; CURRENT APPLICATION NUMBER: US/09/602,877A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 251
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-602-877A-93

```

RESULT 45
US-09-071-710-5
; Sequence 5, Application US/09071710
; Patent No. 6130043
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GORDON, JULIAN
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: HODGES, STEVEN C.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: KRATOCHVIL, JON D.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; FOR DETECTING DISEASES OF THE PROSTATE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071.710
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/850.713
; FILING DATE: 02-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6083.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-071-710-5

Query Match 6.5%; Score 220; DB 3; Length 231;
Best Local Similarity 99.6%; Pred. No. 5.9e-35;
Matches 231; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1767 TGGATAGTGCCTCTCTGCTGCCAGGTGGCCCATCCCTGTTATGGGCTCCATTGTCC 1826
DB 1 TGGATAGTGCCTCTCTGCTGCCAGGTGGCCCATCCCTGTTATGGGCTCCATTGTCC 60
QY 1827 AGCTCAGCCAGTCTGCTACTGCCCTATATGTTGTCGCCAGGCTGGGCTGTGGTCGCCA 1886
DB 61 AGCTCAGCCAGTCTGCTACTGCCCTATATGTTGTCGCCAGG-CTGGGTCTGGTCGCCA 119
QY 1887 TTTACTTTGCTACACAGGTAGTATTTGACACAGAGGACTTGGCCAAATACTCAGCGTAGA 1946
DB 120 TTTACTTTGCTACACAGGTAGTATTTGACACAGAGGACTTGGCCAAATACTCAGCGTAGA 179
QY 1947 AAACCTCCAGCACATTGGGGTGGAGGGCTGCCTCACTGGGTGCCAGCTCCC 1998

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 9, 2003, 19:47:29 ; Search time 36 seconds
(without alignments)
7857.460 Million cell updates/sec

Title: US-09-759-143-110
Perfect score: 6418
Sequence: 1 gggaaccagcgtcgacgcgc.....aaaaaaaaaaaaaaaaaaaaa 3410

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09759143/runat_05062003_111319_23042/app_query.fasta_1.3591
-DB=SwissProt_40 -OPMT=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09759143@cgn_1.1.37@runat_05062003_111319_23042 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	496	7.7	530	1 MATP_HUMAN	Q9umx9 homo sapien
2	489.5	7.6	530	1 MATP_MOUSE	P58355 mus musculus
3	398.5	6.4	660	1 YHL1_EBV	P03181 Epstein-bar
4	373	6.0	1453	1 CA11_MOUSE	P11087 mus musculus
5	358.5	5.8	1464	1 CA13_MOUSE	P08121 mus musculus
6	352.5	5.7	1262	1 CA13_CHICK	P12105 gallus gall
7	351.5	5.6	1466	1 CA13_HUMAN	P02461 homo sapien
8	350	5.6	1460	1 CA11_CANFA	Q9xsj7 canis fami
9	348	5.6	1690	1 CA44_HUMAN	P53420 homo sapien
10	344.5	5.5	1049	1 CA13_BOVIN	P04258 bos taurus
11	344	5.5	1453	1 CA11_CHICK	P02452 gallus gall
12	343	5.5	1464	1 CA11_HUMAN	P02452 homo sapien
13	337.5	5.4	1838	1 CA15_HUMAN	P02098 homo sapien
14	330	5.3	1603	1 CA1F_HUMAN	Q07092 homo sapien
15	327	5.1	525	1 STP_SPTOL	Q03411 spinacia ol
16	326	5.2	1459	1 CA12_MOUSE	P28481 mus musculus
17	324.5	5.2	1355	1 CA21_RANCA	O42350 rana capesh
18	324.5	5.2	1418	1 CA12_HUMAN	P02458 homo sapien

C 19	324.5	5.2	1496	1	CA25_HUMAN	P05997 homo sapien
C 20	323	5.0	1685	1	CA54_HUMAN	P29400 homo sapien
C 21	321	5.2	779	1	CA11_BOVIN	P02453 bos taurus
C 22	321	5.2	1366	1	CA21_CANFA	O46392 canis fami
C 23	314.5	5.1	2944	1	CA17_HUMAN	Q02388 homo sapien
C 24	312.5	4.9	1763	1	CA24_ASCSU	P27393 ascaris suu
C 25	312	4.9	1496	1	CA25_HUMAN	P05997 homo sapien
C 26	311.5	5.0	1362	1	CA21_CHICK	P02467 gallus gall
C 27	311.5	4.9	1516	1	CA1H_HUMAN	P39060 homo sapien
C 28	310	5.0	671	1	CA11_RAT	P02454 rattus norv
C 29	309	5.0	1364	1	CA21_BOVIN	P02465 bos taurus
C 30	308	4.8	1049	1	CA13_BOVIN	P04258 bos taurus
C 31	308	4.9	1356	1	CA21_ONCMY	O93484 oncorhynch
C 32	307	4.9	1650	1	CA2B_MOUSE	Q64739 mus musculu
C 33	307	4.9	1685	1	CA54_HUMAN	P29400 homo sapien
C 34	306.5	4.9	680	1	CA1A_HUMAN	Q03692 homo sapien
C 35	304.5	4.7	680	1	CA1A_HUMAN	Q03692 homo sapien
C 36	304	4.7	1712	1	CA24_HUMAN	P08572 homo sapien
C 37	304	4.9	1806	1	CA1B_HUMAN	P12107 homo sapien
C 38	303.5	4.7	1758	1	CA24_CAEEL	P17140 caenorhabdi
C 39	302	4.9	1372	1	CA21_RAT	P02466 rattus norv
C 40	301.5	4.8	1372	1	CA21_MOUSE	Q01149 mus musculu
C 41	301	4.8	1736	1	CA2B_HUMAN	P13942 homo sapien
C 42	300.5	4.8	1670	1	CA34_HUMAN	Q01955 homo sapien
C 43	300.5	4.8	1804	1	CA1B_MOUSE	Q61245 mus musculu
C 44	299.5	4.7	680	1	CA1A_MOUSE	Q05306 mus musculu
C 45	299.5	4.8	1516	1	CA1H_HUMAN	P39060 homo sapien

ALIGNMENTS

RESULT 1

MATP_HUMAN
ID MATP_HUMAN STANDARD: PRT; 530 AA.
AC Q9UMX9; Q9BMT3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane-associated transporter protein (AIM-1 protein) (Melanoma antigen AIM1).
DE antigein AIM1).
GN MATP OR AIM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homn.
OX NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A. (ISOFORM AIM-1A).
RC TISSUE=Melanoma;
RX MEDLINE=21115844; PubMed=11221837;
RA Harada M., Li Y.F., El-Gamil M., Rosenberg S.A., Robbins P.F.;
RT "Use of an in vitro immunoselected tumor line to identify shared melanoma antigens recognized by HLA-A*0201-restricted T cells.";
RL Cancer Res. 61:1089-1094(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM AIM-1B).
RC TISSUE=Skin;
RA Strausberg R.;
RN [3]
RP Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
ALTERNATIVE SPLICING.
RA Ferro S.;
RN [4]
RP Unpublished observations (NOV-2001).
RN [4]
RP DISEASE, AND VARIANT LEU-374.
RX MEDLINE=21473748; PubMed=11574907;
RA Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davisson M.T., King R.A., Brilliant M.H.;
RT "Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4.";
RL Am. J. Hum. Genet. 69:981-988(2001).
CC -!- FUNCTION: Melanocyte differentiation antigen. May transport substances required for melanin biosynthesis (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By

Db 370 yCysTrpGlyPheCysIleAsnSerValPheSerSerLeuTyrSerTyrPheGlnLysVa 390
 QY 1315 GCTGTCGACGATTCGGCACTCGACAGTCTATTGGCCAGTGGCAGCTTCCTGT 1374
 Db 390 lLeuValSerTyrIleGlyLeuLysGlyLeuTyrPheThrGlyTyrLeuLeuPheGlyLe 410
 QY 1375 GGCTGCGGGTCCACATGCTGCTCCACAGTGGCGGTGGTGCACAGCTTCAGCCGCCCT 1434
 Db 410 uGlyThrGlyPheIleGlyLeuPheProAsnValTyrSerThrLeuValLeuCysSerLe 430
 QY 1435 CACGGGTTCACCTTCACGCTCGAGATCCCTACACACTGGCTCCCTCCCTACCA 1494
 Db 430 upheGlyValMetSerThrLeuTyrThrValProPheAsnLeuIleThrGluTyrHi 450
 QY 1495 CCGGGAAGACAGTGTCTGCTCCCAATACCGAGGACACTGGAGTGTGCTAGCAGTGA 1554
 Db 450 sArgGluGluGlu-----LysGluArgGlnGlnAla----- 460
 QY 1555 GGACAGCTGATGACCACTTCCTCGCAGGCCCTAAGCTGGAGCTCCCTTCCTAATGG 1614
 Db 461 -----ProGlyGlyAspPro----- 465
 QY 1615 ACAGTGGGTGGAGGAGTGGCTGCTCCACCTCCACCGCGCTCTGCGGGGCCCTC 1674
 Db 465 ----- 465
 QY 1675 TGCTGTGATGTCTCGTACGTGTGGTGGTGGAGCCACCGAGGCCAGGTGTTCC 1734
 Db 466 -----AspAsnSerValArg----- 470
 QY 1735 GGCGCGGGCATGCTGCTGGAGCTCCCATCCATCCCTGGATAGTGTCTGCTCCAGGT 1794
 Db 471 -GlyLysGly-----MetAspCysAlaThrLeuThrCysMetValGlnLeuAlaGlnI 488
 QY 1795 GGCCCCATCCCTTTATGGCTCCATGCTCCAGCTCAGCTCCAGTGTGCTAGCTATAT 1854
 Db 488 eLeuValGlyGlyLeuGlyPheLeuValAsnThrAlaGlyThrValValValVa 508
 QY 1855 GGTGTCTGCCGAGCGCTGGGTCTGGTCCGCTTTACTTTGTCTACACAGGTAGTA 1909
 Db 508 lIleThrAlaSerAlaValAlaLeuIleGlyCysCysPheValAlaLeuPheVal 526
 RESULT 2
 MATP_MOUSE
 ID MATP_MOUSE STANDARD; PRT; 530 AA.
 AC P58355;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Membrane-associated transporter protein (AIM-1 protein) (Melanoma
 DE antigen AIM1) (Underwhite protein).
 GN MATP OR AIM1 OR UW.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ddy; TISSUE=Eye, Kidney, and Uterus;
 RX MEDLINE=21372467; PubMed=11479596;
 RA Fukumachi S., Shimada A., Shima A.;
 RT "Mutations in the gene encoding B, a novel transporter protein, reduce
 RL melanin content in medaka";
 RN Nat. Genet. 28:381-385(2001).
 [2]
 RP SEQUENCE FROM N.A., AND VARIANTS UW-DBR ASN-153 AND PRO-435.
 RX MEDLINE=21473748; PubMed=11574907;
 RA Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davisson M.T.,
 RA King R.A., Brillant M.H.;
 RT "Mutations in the human orthologue of the mouse underwhite gene (uw)
 RT underlie a new form of oculocutaneous albinism, OCA4.";
 RL Am. J. Hum. Genet. 69:981-988(2001).
 CC -!- FUNCTION: Melanocyte differentiation antigen. May transport

CC substances required for melanin biosynthesis (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By
 CC similarity).
 CC -!- TISSUE SPECIFICITY: Melanocytes, eyes, kidney and uterus.
 CC -!- DISEASE: Defects in MATP are the cause of the UW-dbr phenotype
 CC that results in loss of nearly all pigmentation in the homozygous
 CC state.
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF360357; AAK81713.1; -
 CC MGD; MG1:2153040; Matp;
 CC Melanin biosynthesis; Transmembrane; Antigen; Glycoprotein; Vision;
 CC Disease mutation; Albinism.
 CC
 CC DOMAIN 1 45
 CC TRANSMEM 46 56 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 67 68 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 69 89
 CC DOMAIN 90 105
 CC TRANSMEM 106 126
 CC DOMAIN 127 138
 CC TRANSMEM 139 159
 CC DOMAIN 160 184
 CC TRANSMEM 185 205
 CC DOMAIN 206 216
 CC TRANSMEM 217 237
 CC DOMAIN 238 318
 CC TRANSMEM 319 339
 CC DOMAIN 340 366
 CC TRANSMEM 367 387
 CC DOMAIN 388 398
 CC TRANSMEM 399 419
 CC DOMAIN 420 425
 CC TRANSMEM 426 446
 CC DOMAIN 447 477
 CC TRANSMEM 478 498
 CC DOMAIN 499 504
 CC TRANSMEM 505 525
 CC DOMAIN 526 530
 CC CARBOHYD 536 556
 CC VARIANT 153 153
 CC VARIANT 435 435
 CC SEQUENCE 530 AA; 57961 MW; F4EDED07916D9FC CRC64;
 SQ
 Alignment Scores:
 Pred. No.: 3.27e-15 Length: 530
 Score: 489.50 Matches: 146
 Percent Similarity: 41.25% Conservative: 85
 Best Local Similarity: 26.07% Mismatches: 228
 Query Match: 7.63% Indels: 101
 DB: 1 Gaps: 12
 US-09-759-143-110 (1-3410) x MATP_MOUSE (1-530)
 QY 332 CAGCTCTTGGTGGTCAACCTTTGGCTGGAGGTGTGTTGGCCGAGGCATC 391
 Db 34 ArgLeuValMetHisSerMetAlaMetPheGlyArgGluPheCysTyrAlaValGluAla 53
 QY 392 ACCTATGTGGCGCTCTGCTGCTGGAAGTGGGGTAGAGGAGTTCATGACCATGGTG 451
 Db 54 AlaTyrValThrProValLeuLeuSerValGlyLeuProLysSerLeuTyrSerMetVal 73
 QY 452 CTGGGCATTTGGTCCAGTGTGGCGCTGTGTGTCCTCCCTCCTAGCTCAGCAGTGAC 511
 Db 74 TrpLeuLeuSerProIleLeuGlyPheLeuLeuGlnProValValGlySerAlaSerAsp 93

QY		512	CAC	TGGCGTGGAGC	GCTATG	GC	CCGCCGCGCCCTTCAT	CTGGCACATGCTTGCGCATC	571	
Db		94	HIS	CysArgAla	argTrpGlyArg	ArgArg	prgTrpIleLeu	IleMetMet	113	
QY		572	CTC	TGAGCCTCTTT	CTATCCC	AAGGCGGTG	CTAGCAGGCGTGTG	CCCGGAT	631	
Db		114	Leu	GlyMetAla	LeuTyrl	LeuAsnGly	AspAlaVal	SerAlaLeuValAla	Asn	133
QY		632	CCC	AGGCC-----	CTGGAGCTGG	CAC	TGCTCATCTGGCGGTGGGCTGG	CATG	682	
Db		134	Pro	ArgGlnLys	LeuIleTrpAla	IleSerIleThr	MetValGlyVal	ValLeuPheAsp	153	
QY		683	TTT	GTGGCCAGGTG	CTCTCACTCC	ACATGGAGGCC	TGCTCTCTGACCTCTTCC	CGGAC	742	
Db		154	Phe	SerAlaAsp	PheIleAspGly	ProIleLysAla	TrpLeuPheAspValCys	SerHis	173	
QY		743	CCG	ACACATGTG	CGCCAGGCTACT	CTGTCTATG	CCCTCATGATCATGCTTGGGGCTGC	802		
Db		174	Gln	AspLys--	-GluLysGly	LeuHisIstyr	HisAlaLeuPheThrGly	PheGlyAla	192	
QY		803	CTG	GGCTACTCTG	CTGCTG	CATTGACTGG	GACACAGTCCCTGGCCCCCTACTAC	TGGGC	862	
Db		193	Leu	GlyTrpIleLeu	GlyAlaIleAsp	TrpValHisLeu	AspLeuGlyArgLeu	LeuGly	212	
QY		863	ACC	GAGGAGTGC	CTTTGGCTG	CTCACCTCAT	CTCTCTCACTCTGCTAGCAGCC	922		
Db		213	Thr	GluPheGln	ValMetPhe	PheSerAlaLeu	ValLeuIleLeuCysPhe	IleThr	232	
QY		923	ACA	CTGCTGTG	TGCTGAGGAGG	CAGCGCTG-----	GGCCCCACGAGCCA-----	967		
Db		233	His	LeuCysSeri	LeProGluAla	ProLeuArgAsp	AlaIleThrAspPro	SerGln	252	
QY		968	-----	-GCAGAAGG	CTGTGCGGCCCTCT	CTGTGCGGCCCACTGCTG	CCATGC	1015		
Db		253	Gln	AspProGln	GlySerSerLeu	SerAlaSerGly	MethHisGlu	Trp-----	268	
QY		1016	CGG	CGCGCTGG	CTTTCCGGAAC	CTGGCGCC-----	-----	1048		
Db		269	Gly	SeriLeu	LysValLysAsn	GlyAlaAsp	ThrGluGlnPro	ValGlnGlu	Trp	288
QY		1049	-----	-----	-----	-----	-CTGCTCCCCGGCTGC	CACCAG	1069	
Db		289	Lys	AsnLysLys	ProSerGly	GlnSerGlnArg	ThrMetSerMetLys	SerLeuLeuArg	308	
QY		1070	CTG	TGCTGCGCATG	CGCCCGCCAC	CTCGCGCGCTCTTCGTGCTG	CTGAGCTGCAGCTGG	1129		
Db		309	Ala	LeuValAsn	MetProSer	HisTsrArgCys	LeuLysValSerHis	IleLeuGly	Trp	328
QY		1130	ATG	CAC	TATCATG	CTTACCGTGT	TTTACAGGATTTCTGGCGGAGGGCTG	TACCAG	1189	
Db		329	Thr	AlaPheLeu	SerAsnMetLeu	PhePheThrAsp	PheMetGlyGlnIle	ValTyrHis	348	
QY		1190	GGC	CTGCCAGAG	CTGAGCGCGG	CACCCAGCGCGGAGAC	ACTATGATGAAGCGTTCGG	1249		
Db		349	Gly	AspProTrp	GlyAlaHis	AsnSerThrGlu	PheLeulleTyrGluArg	GlyValGlu	368	
QY		1250	ATG	GGCAGCCTG	GGGCTGTTC	TGTCAGTGGCGCATCTCC	TGCTGCTTCTCTCTG	GTGTCATG	1309	
Db		369	Val	GlyCysTrp	GlyLeuCysIle	AsnSerValPhe	SerSerValTyrSer	TyrPheGln	388	
QY		1310	GAC	CGCTGTG	CAGCATTCGG	CACATCGAGCAGTCTAT	TGGCCAGTGTGGCAGCTTTC	1369		
Db		389	Lys	AlaMetVal	SerTyrIle	GlyLeuLysGlyLeu	TyrPheMetGly	TyrLeuLeuPhe	408	
QY		1370	CCT	GTGCTGCG	GTGCCATG	CTCTGCCACAGTGTGGCGGTG	GTGACAGCTTTCAGCC	1429		
Db		409	Gly	LeuGlyThr	GlyPheIle	GlyLeuPhePro	AsnValTyrSerThr	LeuValLeu	Cys	428
QY		1430	GCC	CTCACC	GGGTTTCACTTCT	CTCAGCCCTGCAGATCTG	CGCTTACACACTGG	CGCTCCCTC	1489	
Db		429	Ser	MetPheGly	ValMetSerSer	ThrLeuTyrThr	ValProPheAsn	IleLeuIle	Glut	448
QY		1490	TAC	CACCGGGA	AGACAGGTG	TCTCCTGCCCAAATAC	CGAGGGACACTGGAGGTG	TGATGC	1549	

Db	449	TyrHisArgGluGluGlu			---	LysGluGlyGlnGluAla	-----	461
Qy	1350	AGTGAGGACAGCCTGATGACACAGCTTCTGCCAGGCCCTAAGCGCTGGAGCTCCCTTCCT			---	ProGlyGlyProAspAsn	467	
Db	462	-----			---	ProGlyGlyProAspAsn	467	
Qy	1610	AATGGACAGTGGTGTCTGGAGGACAGTGGCTGCTCCACCTCCACCCGGCTCTGCGGG			---	AspCysAla	477	
Db	468	GlnGlyArg			---	AspCysAla	477	
Qy	1670	GCCCTCTGCCTGATGTCTCCGTACGTGTGGTGGGTGAGCCACCGAGGCGAGGTG			---	AlaGlnIle	488	
Db	478	AlaLeuThrCysMetValGlnLeu			---	AlaGlnIle	488	
Qy	1730	GTTCCGGGGCGGGCGATCTGCTCGACTCGCCATCTGATAGTGCCTTCTCTGCTCTCC			---	AlaGlnIle	488	
Db	489	LeuValGlyGlyGly			---	AlaGlnIle	488	
Qy	1790	CAGGTGCCCCATCCCTGTTTATGGTTCATTTGTCAGCTTCAGCTTCAGCTTCAGTGC			---	AlaGlnIle	488	
Db	494	-----			---	AlaGlnIle	488	
Qy	1850	TATATGTGTCTGCCGAGCGCTGGGTCTGGTCCGCCATTTACTTTGTGTACACAGGTAGTA			---	AlaGlnIle	488	
Db	507	ValValIleThrAlaSerAlaValSerLeuIleGlyCysPheValAlaLeuPheVal			---	AlaGlnIle	488	
RESULT 3								
YHLI_EBV	ID	YHLI_EBV	STANDARD;	PRT;	660	AA.		
AC	PO3181;							
DT	21-JUL-1986 (Rel. 01, Created)							
DT	21-JUL-1986 (Rel. 01, Last sequence update)							
DT	15-JUL-1998 (Rel. 36, Last annotation update)							
DE	Hypothetical BHUF1 protein.							
OS	Epstein-barr virus (strain B95-8) (Human herpesvirus 4).							
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;							
OC	Gammapherpesvirinae; Lymphocryptovirus.							
OX	NCBI_TaxID=10377;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RX	MEDLINE=84270667; PubMed=6087149;							
RA	Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,							
RA	Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Sequin C.,							
RA	Tufnell P.S., Watfoll B.G.;							
RT	"DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";							
RL	Nature 310:207-211(1984).							
CC	-----							
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration							
CC	between the Swiss Institute of Bioinformatics and the "EMBL outstation"							
CC	the European Bioinformatics Institute. There are no restrictions on its							
CC	use by non-profit institutions as long as its content is in no way							
CC	modified and this statement is not removed. Usage by and for commercial							
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ ;							
CC	or send an email to license@isb-sib.ch).							
CC	-----							
DR	EMBL; V01555; --; NOT_ANNOTATED_CDS.							
DR	PIR; A03742; Q0BE3.							
KW	Hypothetical protein; Early protein; Repeat.							
FT	DOMAIN 149 648 4 X 125 AA TANDem REPEATS.							
FT	REPEAT 149 273 1.							
FT	REPEAT 274 398 2.							
FT	REPEAT 399 523 3.							
FT	REPEAT 524 648 4.							
SQ	SEQUENCE 660 AA; 66244 MW; 86DAID67A37152A2 CRC64;							
Alignment Scores:								
Pred. No.:		3,46e-11	Length:	660				
Score:		398.50	Matches:	237				
Percent Similarity:		21.86%	Conservative:	37				
Best Local Similarity:		27.56%	Mismatches:	276				
Query Match:		6.40%	Indels:	311				
DB:		1	Gaps:	51				


```

QY 425 -----CCCCCACTTCCAGCAGCAGCAGCGGCGGCACATAGGTGATGCTGCGGCCAAACAC 373
Db      ||||| ||| ||| |||||
QY 542 gGlyHisProProGlyAlaGlyGln-----ArgProSerG1 555
QY 372 ACCTCCAGCCCAAGGTTAGCAGTTGACCCAGCAGAGCTGGGCTTCGCGTCCGCGAC 313
Db      ||| ||| ||| |||||
QY 555 yProThrGlyGlyArg-----ProAlaAlaProGly-----Al 566
QY 312 AGCGGCTCACCACACAGCTCTGGCATTAGTGGCCAGCGGGTAGGCTCAGGGGCC 253
Db      ||||| ||| ||||| ||| |||||
QY 566 aProGlyThrProAlaAlaProGlyPro-GlyGlyAlaAlaValProSerGly----- 584
QY 252 GTTCAGGCTCAGCAAGCTCTGCTGCTGCTGCTGCTCCAGAGCTCGGCGCTCTCTC 193
Db      ||||| ||| ||||| ||| |||||
QY 585 -----AlaThrProHisProGluArg---GlySerGlyPro-----AlaAspProp 599
QY 192 CITGCTGCCGCACTGCTAGCATCAGCAGCGGCCCAATTCGCCAGCCCTTCCTC 133
Db      ||| ||||| ||| ||||| ||| |||||
QY 599 roAlaAlaAlaArgLeuProGluArg-GlnGluProArgLeuProGluAla 618
QY 132 CGGCTCAGCTCTCAGCCCTGCTCAACACCTGCTGTGGGCGACCTCAGTGGGAC 73
Db      ||| ||| ||| |||
QY 619 AlaAlaGln---ArgCysProAlaGlyProProThrArgSerGlyAlaAlaGln 637
QY 72 AGCTCATCATCATGATCTCTGGC-----CGAGCGCGCGCTGTCTACCCCGGA 25
Db      ||||| ||||| ||| ||||| ||| |||||
QY 638 ArgThrHisArgArgProProGlyCysProArgSerAlaArg---AsnProGly 654
RESULT 4
CALL_MOUSE
ID CALL_MOUSE STANDARD; PRT: 1453 AA.
AC P11087; Q60635;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1 OR COL1A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RX MEDLINE=96033240; PubMed=8535610;
RA Li S.W., Khillan J., Prockop D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
of type I procollagen.";
RL Matrix Biol. 14:593-595(1995).
RN [2]
RX SEQUENCE OF 518-1128 FROM N.A.
RX MEDLINE=86137403; PubMed=3841523;
RA French B.T., Lee W.-H., Maul G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
collagen protein.";
RL Gene 39:311-312(1985).
RN [3]
RX SEQUENCE OF 735-1130 FROM N.A.
RX MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1(I) procollagen gene:
evidence for a mouse B1 element within the gene.";
RL Mol. Cell. Biol. 2:1362-1371(1982).
RN [4]
RX SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX MEDLINE=83157109; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences.";
RL DNA 1:59-69(1981).
RN [5]
RX SEQUENCE OF 1442-1453 FROM N.A.

```

```

RX MEDLINE=88124276; PubMed=3340560;
RA Mooslehner K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
of the 3'-untranslated region.";
RL Nucleic Acids Res. 16:773-773(1988).
CC -|- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
(FIBRILLAR FORMING COLLAGEN).
CC -|- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -|- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
HYDROXYAPATITE.
CC -|- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -|- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U08020; AAA88912.1; -
DR EMBL: X15896; CAA33904.1; -
DR EMBL: M14423; AAA37333.1; -
DR EMBL: M17491; AAA37334.1; -
DR EMBL: X06753; CAA29927.1; -
DR EMBL: K03036; AAA37332.1; -
DR EMBL: K03029; AAA37332.1; JOINED.
DR EMBL: K03030; AAA37332.1; JOINED.
DR EMBL: K03031; AAA37332.1; JOINED.
DR EMBL: K03032; AAA37332.1; JOINED.
DR EMBL: K03033; AAA37332.1; JOINED.
DR EMBL: K03034; AAA37332.1; JOINED.
DR EMBL: K03035; AAA37332.1; JOINED.
DR PIR: A23982; A23982.
DR MGI: 88467; Collal.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR001007; VWFC.
DR Pfam: PF01391; Collagen; 18.
DR Pfam: PF01410; COLFI; 1.
DR ProDom: PD000007; Collagen; 1.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00214; VWFC; 1.
DR PROSITE: PS01208; VWFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 29 87 VWFC.
FT DOMAIN 152 167 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 168 1181 TRIPLE-HELICAL REGION.
FT DOMAIN 1182 1207 NONHELICAL REGION (C-TERMINAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1354 1354 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 734 736 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1082 1084 CELL ATTACHMENT SITE (POTENTIAL).
FT CONFLICT 1450 1450 A -> V (IN REF. 5).
SQ SEQUENCE 1453 AA; 137944 MW; 38802E535DF81808 CRC64;
Alignment Scores:
Pred. No.: 4,31e-10 Length: 1453
Score: 373.00 Matches: 285
Percent Similarity: 31.98% Conservatives: 47
Best Local Similarity: 27.46% Mismatches: 371
Query Watch: 5.99% Indels: 336
DB: 1 Gaps: 55

```

QY	1674	GAGCC---CCGCAGACCGCGGTGGAGGT-----GGGACGAGG-----CCA	1633
Db	343	GLUALaglyProGlnGlyAlaAargglySerGluGlyProGlnGlyValAargGlyGluPro	362
QY	1635	CTGCCTCCAGCACCCACGTCTCCATTAGGG---AAGGAGCTCCAGCGTTAGG---	1585
Db	363	GlyProProGlyProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGln	382
QY	1584	CTGGGAGGAAAGCTGTGTCATCAGGCTGTCTCTACTGCTACTAGCACCTCCAGTGTCCCTCGG	1525
Db	383	ProGlyAlaLys-GlyAlaAsnGlyAlaPro-----	393
QY	1524	TATTGGGAGGAAACACCTCTCTCCGTGTGTAGGAGGAGCCAGTGTGTAGGCGAGG	1465
Db	393	Ylle---AlaGlyAlaProGlyPheProGlyAlaAarggly-----ProSerGI	408
QY	1464	ATCTGAGGCTGAGAAAGGTGAACCGCGGTGAGGCGCGCTGAAGCTGTCACACCGGCACAC	1405
Db	408	YPro-GlnGlyProSerGlyProProGlyProLysGlyAsnSerGlyGluProGly---	427
QY	1404	CTGTGGACAGGATGTGGCACCGGACCGGACCCAGGGAAGACTGCCACACTGGCCCAATAG	1345
Db	427	laProGlyAsnLysGlyAspThrGlyAlaLysGlyGluProGlyAlaThrGly-----	444
QY	1344	ACTGCTCGAGTCGCGAATCCTCCACGACGCGCTCCATGACACAGAGAGAAGACCGGGAG	1285
Db	445	-----ValGlnGlyPro--ProGlyProAlaGlyGluGluGlyLy	457
QY	1284	ATGGCGACATGTCAGGAACA-----GCCCGAGGTGCCCATCCGAAACGC---	1242
Db	457	sArgGlyAlaAargGlyGluProGlyProSerGlyLeuProGlyProGlyGluArgGI	477
QY	1241	-----CTTCATCATAGTGTCTCCGGCGCTCGGTGCCCGC	1207
Db	477	yGlyProGlySerArgGlyPheProGlyAlaAspGlyValAlaGlyProLysGlyProse	497
QY	1206	TCAGCTCTGGGCACGCCCTGGTTACAGCCCTCGCCACGAAATCGTGTAAAAACACGCTG	1147
Db	497	rGlyGluAargGlyAlaProGlyProAlaGlyProLysGlySerPro-----	512
QY	1146	AAGGTCATGATGCCATCCAGCTGCACAGCTCAGCCACAGAGACCGCGCGAGGTGCGG	1087
Db	512	-----	512
QY	1086	GGCATCGGCACACAGCTGTGTGACGCCGGGAAGCAGGCGGCCAGGT-----	1038
Db	513	-----GlyGluAlaGlyArgProGlyGluAlaGlyLeuProGlyAlaLysGlyLe	529
QY	1037	-TCCGAAAGCCAAAGCGGCCCGCATGACACAGCAGTGGGCGGCAAGGAGGGCGGCAC	979
Db	529	uThrGlySerProGlySerProGlyProAsp-----GlyLysThrGlyProProGlyPr	547
QY	978	AGCCCTCTGCTGGCTCGGTGGGGCCCGACGCTGCCT-----	942
Db	547	oAlaGlyGlnAspGlyArgProGlyProAlaGlyProProGlyAlaAargGlyGlnAlaGI	567
QY	941	-----CCTCAGCCACCAACAGTGTGGCTGTGTACGAGGTGAGGAAGATGAGG	895
Db	567	yValMetGlyPheProGlyProLysGlyThrAla-----GlyGluProGI	582
QY	894	GTGAGCAGGCCAAAGAGGACATCCT-----	870
Db	582	YlysAlaGlyGluAargGlyLeuProGlyProProGlyAlaValGlyProAlaGlyLysAs	602
QY	869	-----CCTGGGTGCCAGGT---AGGGGGCCAGGSCACTGGTGTGCCAG	829
Db	602	pGlyGluAlaGlyAlaGlnGlyAlaProGlyProAlaGlyProAlaGlyGluAargGlyGI	622
QY	828	TCAATGGCAGGAGGA-----GGTAGCCACAGGACGCCCA	793
Db	622	uGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyProAlaGlyProProGI	642

Db 464 lYlYs-AspGlySerProGlyGluPro--GlyAlaAsnGlyLeuProGlyAlaAlaGlyG 483
QY 1995 AGCTGGGACCCAGT-----GAGGCG 1975
Db 483 luArgGlyProSerGlyPheArgGlyProAlaGlyProAsnGlyIleProGlyGluLysG 503
QY 1974 GCGCTCCA-----CCCCAATGTCTGGAGTATTTCTACGCTGAGTATTG 1930
Db 503 lYProProGlyGluArgGlyGlyProGlyProAlaGlyPro-----ArgGlyValAlaG 521
QY 1929 GCCAGTCCCTCTTCAATACTACTCTGTAGCAAAAGTAATGGCCAGCAGACCCAGG 1870
Db 521 lYcLuProGly-ArgAspGlyThr-Pro-----GlyGlyProGly 533
QY 1869 CTGCGGCACACCATATAGGACAGTGGCTGAGCTGGACACATGGAGCCCAT 1810
Db 533 yIleArgGlyMetProGlySerProGlyGlyProGlyAsnAspGlyLysProGlyProPr 553
QY 1809 ACAGGGATG-----GGGCCACTGGGACAGGAGGAC----- 1773
Db 553 oGlySerGlnGlyGlySerGlyArgProGlyProGlyProSerGlyProArgGlyGly 573
QY 1772 -----TATCCAGGATGGGAGGTCCAGGCAGATGCCCGGCCGGRAC- 1730
Db 573 nProGlyValMetGlyPheProGlyProLysGly-AsnAspGlyAlaProGlyLysAsnG 593
QY 1729 -----CACCTGCTCCGCTGGCTCACCCACACACACACACGTCACGAGACATCA 1681
Db 593 lYcLuArgGlyGlyProGlyGlyProGlyLeuPro----- 604
QY 1680 CAGGACAGAGCCCGCAGAGCCGGTGGAGTGGAGCCAGCCACTGCTCCAGCACCC 1621
Db 605 -----GlyProAlaGlyLys-AsnGlyGluThrGlyProGlnGlyProProGlyPro 621
QY 1620 ACCTGCTCCATTAGG-----AAGGGA 1600
Db 622 ThrGlyProAlaGlyAspLysGlyAspSerGlyProProGlyProGlnGlyLeuGlnGly 641
QY 1599 GTCCAGGCTTAGG----- 1585
Db 642 lIleProGlyThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLys 661
QY 1584 -----CCTGGCAGGAAGCTGGTCATCAGCTGCTCCTCACTG 1549
Db 662 GlyGluValGlyAlaProGlyAlaProGly-GlyLysGlyAspSerGlyAlaProGlyGly 681
QY 1548 CTAGCACCTCCAGTGTCCCTCGTATT-----TGGCAGGAACACCTGC 1504
Db 681 uArgGlyProProGlyThrAlaGlyIleProGlyAlaArgGlyGlyAlaGlyProProGly 701
QY 1503 TTCTCCCGTGTAGAGG-----AGGCCAGTGTGTAGGCGAGGATGTC 1459
Db 701 yProGluGlyGlyLysProAlaGlyProProGlyProProGlyAlaSerGlySerPr 721
QY 1458 AGGGCTGAGAAGGTGNACCGGTGAGCGGTGAAGCTGTACCCAGCCACCTGTCG 1399
Db 721 oGlyLeu-GlnGlyMetProGlyGluArgGly----- 731
QY 1398 GACAGGCATGTGCGACCGGCACACAGGAAAGCTGCCACACTGGCCAAATAGACTGCT 1339
Db 732 -----GlyProGlySerProGlyProLysGlyGlyGluProGlyGlyAla 748
QY 1338 CGAGTCCGAATCGCTGCA-----CCAGCCGGT----- 1311
Db 748 laGlyAlaAspGlyValProGlyLysAspGlyProArgGlyProAlaGlyProIleGlyP 768
QY 1310 --CCATGACACAGAGACACAGCAGATGCGGCACTGCAGGACACACCCAGCCTGC- 1254
Db 768 roProGlyProAlaGlyGlnProGlyAsp---LysGlyGluGlySerProGlyLeup 787
QY 1253 -----CCATCGGAACGCTTCATCATAGTGTCTCCGGGC 1220

Db 787 roGlyIleAlaGlyProArgGlyGlyProGlyGluArgGlyGluHis-----GlyP 804
QY 1219 CTCGGTCCCGGCTCAGCTCTGGCAGCCCTGTGTACAGCCCTCGCCACCAAAATCCGT 1160
Db 804 roProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGluPro----- 820
QY 1159 GTAAACAGCGTGAAGTGCATGAGTGCATCCAGCTGCACAGCTCAGCCACGAAGACCG 1100
Db 820 ----- 820
QY 1099 GCGCAGGTCGGGGCATCGGCAGCAGACAGCTGGTCAGCCGGGGAAGCAGCGGCCCA- 1041
Db 821 -----GlyAlaLysGlyGluArgGlyAlaProG 830
QY 1040 -----GGTTCGGAAAGCCAAAGCGGCCGCGCATGGACAGCTGGGGCGACAAGGA 989
Db 830 lYcLuLysGlyGlyGlyProProGlyProAla----- 841
QY 988 GGGGGCCGACAGCCCTTCTGCTGGTCTGGTGGGGCCAGCGCTGCTCTCAGCCACAG 929
Db 842 --GlyProThr-----GlySerSerGlyProAlaGlyProProGlyProGlnG 857
QY 928 CAGTGTGGCTGCTACGACGAGGTGAGGAAGATGAGGTGAGCAGCAGCAAAAGAGCACCTCT- 870
Db 857 lYVal-----LysGlyGluArgGlySerProG 866
QY 869 -----CCTGGTGGCCAGGTAGGGCCAGGCGACTGTGTCCAGTCAATGGCAGCGAG 815
Db 866 lYcLuProGlyThrAlaGlyPheProGlyGlyArgGlyLeuProGlyProProGlyAsnA 886
QY 814 GAGGTAGCCAGCAGCCCGCCAGACTGATCATGAAGGCATACAGAGTAGGCTTGGCG 755
Db 886 snGlyAsnProGlyProProGlyPro-----SerGlyAlaProGlyL 900
QY 754 ACAGTGTGCTCG-----GGTCCCGAAGAGGTGCAGAGACAG 719
Db 900 yAspGlyProProGlyProAlaGlyAsnSerGlySerProGlyAsnProGlyIleAlaG 920
QY 718 GGCCTCAGTGGAGTGAAGCAGCAGCTGGCCACAGAGTCCAGAGCCCGCCAGGAT 659
Db 920 lYProLysGlyAlaGlyGlnProGlyGlyLysGlyPro-----ProGly- 935
QY 658 GAGCAGTCCAGCTCCAGGGGCTGGGATCG-----GSCACAGCAGCCCTCTAGCCAGCC 602
Db 936 --AlaGlnGlyProProGlySerProGlyProLeuGlyIleAlaGlyLeuThrGlyAlaA 955
QY 601 GGCCTTTGGATGAGAAGAGCTCAGCAGGATGCCCAAGCAGACAGTCCAGATGAAGG 542
Db 955 rgGlyLeuAlaGlyProProGlyMetProGly---ProArgGlySerProGlyProGlnG 974
QY 541 CCGCGCGCGCATAGC-----GTCCACCGCAGTGTGTACTGTGCTAGGCTAGGAGCGG 488
Db 974 lYlIleLysGlyGluSerGlyLysProGlyAlaSerGlyHis-----AsnG 989
QY 487 GACACAGA-----CCAGGCCAGCAGTGGACCAATGCCAGCAGCCATGTATGAACIT 434
Db 989 lYcLuArgGlyProProGlyPro-----G 997
QY 433 CTCCTCTACCCCTTCCAGCAGCAGAGGCGGCACATAGGTGATGCTCGCGGCCAACA 374
Db 997 lNcGlyLeuProGlyGlnProGlyThrAlaGly-----GluProGlyAla 1011
QY 373 CACCTCCAGGCCAAAGTTAGCAGGTTGACACAGAGAGCTGGCTTTCGGTGGCC----- 318
Db 1011 rgAspGlyAsnProGlySerAspGlyGlnProGlyArgAspGlySerProGlyGlyLysG 1031
QY 317 -----GCAGCAGGCGCTCACCCAGCAGCTCTGGACCACTAGTGGG--- 278
Db 1031 lYAspArgGlyGluAsnGlySerProGlyAlaProGlyAlaProGly-HisProGlyPro 1050
QY 277 CCAGCGGGTAGGCTCAGGGCGCTTACAGGCACTCCAGAACTGTCTCTCGGCTCT 218
Db 1051 ProGlyProValGlyProSerGlyLysSerGly-----AspArgGlyGlyThrGlyPro 1068

QY 217 GCT---CCAGAAGCTGGCGCTCTCCTCTGTCGCGC-----CAACTG 176
DB 1069 AlAGlyProSerGlyAlaProGlyProAlaGlyAlaArgGlyAlaProGlyProGlyGly 1088
QY 175 CCTAGGAATCAGCCAGCG-----GCCCAFTTC 149
DB 1089 ProArgGlyAspGlyGlyGluThrGlyGluArgGlySerAsnGlyIleGlyHisArg 1108
QY 148 TGCCAGCCCTTGTGGCGGCTCCAGCTCTCAGCCCATGCTCAACACCTCTGCTGTGGG 89
DB 1109 GlyPheProGlyAsnProGlyProGly-Ser-----ProGlyAlaAlaGly 1124
QY 88 GCACCTCAGT-----GGGACACGCTCATCA 62
DB 1124 yHisGlnGlyAlaIleGlySerProGlyProAlaGlyProArgGlyProValGlyProH 1144
QY 61 CTCAGATCCGCGCGA---GGCGCGCGCTGTCCACCGGA 25
DB 1144 sGlyProGlyLysAspGlyThrSerGlyHisProGly 1157

RESULT 6

CA13_CHICK
ID CA13_CHICK STANDARD; PRT: 1262 AA.
AC P12105; P79758; P79759; Q90794; Q92029;
DT 01-OCT-1989 (Rel. 12, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(III) chain precursor (Fragments).
GN COL3A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-886 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94266842; PubMed=8206952;
RA Nah H.-D., Niu Z., Adams S.L.;
RT "An alternative transcript of the chick type III collagen gene that
RT does not encode type III collagen."
RL J. Biol. Chem. 269:16443-16448(1994).
RN [2]
RP SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.
RX MEDLINE=84270696; PubMed=654770;
RA Yamada Y., Liau G., Mudryj M., Ohici S., de Crombrughe B.;
RT "Conservation of the sizes for one but not another class of exons in
RT two chick collagen genes."
RL Nature 310:333-337(1984).
RN [3]
RP SEQUENCE OF 977-1262 FROM N.A.
RX MEDLINE=83220816; PubMed=6856474;
RA Yamada Y., Kuhn K., de Crombrughe B.;
RT "A conserved nucleotide sequence, coding for a segment of the C-
RT propeptide, is found at the same location in different collagen
RT genes."
RL Nucleic Acids Res. 11:2733-2744(1983).
CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC ALONG WITH TYPE I COLLAGEN.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

CC EMBL; U07973; AA83407.1; -
DR EMBL; X00822; CAB52686.1; -
DR EMBL; X00823; CAB52686.1; JOINED.
DR EMBL; X00826; CAA25397.1; ALT_SEQ.
DR EMBL; X00825; CAA25397.1; JOINED.
DR EMBL; X00827; CAA25398.1; -
DR EMBL; X00828; CAA25399.1; -
DR EMBL; X00830; CAA25401.1; -
DR EMBL; X00831; CAA25402.1; -
DR EMBL; K02302; AAD15299.1; -
DR EMBL; K02301; AAD15298.1; -
DR EMBL; M36662; AAA18519.1; ALT_SEQ.
DR PIR; A05269; A05269.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWFC.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWFC; 1.
DR PROSITE; PS01208; VWFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 144 AMINO-TERMINAL PROPEPTIDE (BY
FT CHAIN 145 1003 SIMILARITY).
FT PROPEP 1004 1262 COLLAGEN ALPHA 1(III) CHAIN.
FT DOMAIN 29 88 CARBOXYL-TERMINAL PROPEPTIDE (BY
FT DOMAIN 145 164 SIMILARITY).
FT DOMAIN 165 994 NONHELIICAL REGION (N-TERMINAL) (BY
FT DOMAIN 995 1003 TRIPLE-HELICAL REGION (BY SIMILARITY).
FT NON_CONS 886 887 NONHELIICAL REGION (C-TERMINAL) (BY
FT NON_CONS 922 923 SIMILARITY).
FT DISULFID 994 994 INTERCHAIN (BY SIMILARITY).
FT DISULFID 995 995 INTERCHAIN (BY SIMILARITY).
FT MOD_RES 262 282 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 859 859 HYDROXYLATION (BY SIMILARITY).
FT CARBOHYD 1163 1163 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 96 96 E -> K (IN REF. 2).
FT CONFLICT 1132 1132 F -> S (IN REF. 3).
SQ SEQUENCE 1262 AA; 121249 MW; 96ABE7B2E9DEB43D CRC64;
Alignment Scores:
Pred. No.: 3.55e-09 Length: 1262
Score: 352.50 Matches: 277
Percent Similarity: 31.42% Conservative: 51
Best Local Similarity: 26.53% Mismatches: 353
Query Match: 5.66% Indels: 364
DB: 1 Gaps: 58
US-09-759-143-110 (1-3410) x CAL3_CHICK (1-1262)
QY 2479 GTGTGT---TGCCCTCAGGACTCTCCCTACAATAAGTCATATGTTCAATCCCATG 2423
DB 51 IleCysValCysAspSerGlySerValLeuCysAspAspIleCysAspAspGlnGlu 70
QY 2422 GAGGAGTGTTCATCTCAGAACTCCCATGCAAGAGCTACATTAACGAAGTCGACGTT 2363
DB 71 LeuAspCysProAsnProGlnProGluProGluGluCys-----CysProVal 86
QY 2362 AAGGGCTTAGAGATGGGAACAGCTGACTGAGTTTATTC---GCTCCCAAAACCCCT 2306
DB 87 CysProGlnThrThrProGlnProThrGluLeuProThrThrGlnGlyProLys----- 104
QY 2305 TCTCTAGGTGTCTCTCACTAGGAGGCTAGCTGTTAACCCCTGAGCTGGGTATATCA--- 2249
DB 105 -----GlyAspProGly 108

[illegible]

Qy	1322	GCACCGCCGGTCCATGACCA- 	-----GAG 1299
Db	403	---ProGlyGlyProGlyGlyArgGlyLeuProGlyProGlyThrSer 	421
Qy	1298	AGAAGACGGAGATGGCCACTGCAGGAACAGCC- 	1263
Db	422	GlyAsnProGlyAlaLysGlyThrProGlyGluProGlyLysAsnGlyAlaLysGlyAsp 	441
Qy	1262	CCAGGTCGCCATCCGAACGCCTTCATCATAGTGTCTCCGGGCTCGGTCCCGCTCAG 1203 	
Db	442	ProGlyProLysGlyGluArgGlyGluAsnGlyThrProGlyAlaArgGlyProGly 	461
Qy	1202	CTCTGGGCAGCCCTGGTACAGCCCTCGCCACGAATCCGTGTAACACAGCGTGAAGG 1143 	
Db	462	GluGluGlyArgGlyAlaAsnGlyGluProGly- 	474
Qy	1142	TCATGAGTCATCCAGCTCCACAGCTCACCCACGAAGAGCCGCCA- 	1095
Db	475	AsnGlyValProGlyThr- 	490
Qy	1094	-----GGGTGCGGGCATCGGCAGCACACAGCTGTGCACCGGGGAAGCAGGGCGC 1044 	
Db	491	GlyLeuProGlySerAsnGlyLeuProGlyGluLysGlyProAlaGlyGluArgGlySer 	510
Qy	1043	CCAGGTTCCGAAGCCAGCGGCCCGGATGGACAGCAGTGGGGCAGCAAGAGAGGGG 984 	
Db	511	ProGlyProGlyProSerGlyProAlaGlyAsp- 	524
Qy	983	CCGACAGCCCTTCTGCTGGCTCGGTGGGGCCAGCGCTGCCT- 	942
Db	525	GlnAspGly- 	538
Qy	941	-----CCTCAGCCACAGCA- 	918
Db	539	GlyIleProGlySerProGlySerAspGlyLysProGlyProGlyAsnGlnGlyGlu 	558
Qy	917	CTACGCAGTGAGGAAGATGAGGTGACGAGCCAAAGAGGCACCTCCT- 	870
Db	559	ProGlyArgSerGlyProGlyProAlaGlyProArgGlyGlnProGlyValMetGly 	578
Qy	869	-----CCTGGGTGCCAGGT- 	843
Db	579	PheProGlyProLysGlyAsnGluGlyAlaProGlyLysAsnGlyGluArgGlyProGly 	598
Qy	842	CACGTGGTCCAGTCAATGCAGGCAGGA- 	798
Db	599	GlyProProGlyThrProGlyProAlaGlyLysAsnGlyAspValGlyLeuProGlyPro 	618
Qy	797	CCCCAGACTGATCATGAAGGCATAGACAGATGAGCCTCGGCACAGT- 	750
Db	619	ProGlyProAlaGlyProAlaGlyAspArgGlyGluProGlyProSerGlySerProGly 	638
Qy	749	-----GGTCCGGGTCCCGGAAGAGGTTCAGAGCA- 	720
Db	639	LeuGlnGlyLeuProGlyGlyProGlyProAla- 	657
Qy	719	-----GGGCCTCCAGTGGAGTGAAGC 	699
Db	658	ProGlyProLysGlyAspIleGlyGlyProGlyPheProGlyProLysGlyGluAsnGly 	677
Qy	698	ACACTGGCCACAGAAGTCCAGCAGCCCA- 	663
Db	678	IleProGlyGluArgGlyProGlnGlyProGlyProThrGlyAlaArgGlyGlyPro 	697
Qy	662	GGATGACAGTGCCAGT- 	636
Db	698	Gly- 	714
Qy	635	TGGGATCCGGCAGCAGCCCTCTAGCCAGCGGCCCTTGGATCAGAAAGAGGCTCA 576 	
Db	715	-----ProGlyGlyThrGlyLeuProGlyLeuGlnGlyMetProGlyGluArgGlyAla 	732
Qy	575	GCAGGATGCCCAAGGACAGTCCCCAGATGAAGGGCCGGC- 	522


```

Db      733 SerGlySerProGly-----ProLysGlyAspLysGlyGluProGlyGlyLysGlyAla 750
QY      521 CAC-----GCACTGTCCTACTGCTAGGCTAGGACGGGACACACACAGGC 474
Db      751 AspGlyLeuProGlyAlaArgGlyGluArgGlyAsnValGlyProLeGlyProProGly 770
QY      473 CCAGCACTGACCAATGCCAGCACCATGCTCATGAACCTTCTCTCTA---CCCCCACTT 417
Db      771 ProAla-----GlyProProGlyAspLysGlyGluThrGlyProAlaGlyAlaProGly 788
QY      416 CCAGCACACAGGGCGGCACATAGTGCTGCGCCAAACACACACTCCAGCCCAAGG 357
Db      789 ProAlaGlySerArgGly-----GlyProGlyGluArg 799
QY      356 -----TTAGCAGGTTGACACAGACAGCTGGCTTTCGGTGCC 318
Db      800 GlyGluGlnGlyLeuProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGlu 819
QY      317 GCACGAGGC-----GGCTCACCACCA 297
Db      820 ProGlyGlyLysGlyGluArgGlyProGlyLeuArgGlyGluAlaGlyProProGly 839
QY      296 GCCTCTGGACCATAGTGG-----CCAGCGGGTAGGGCTCAGG 258
Db      840 AlaAlaGlyPro-GlnGlyGlyProGlyAlaProGlyProProGlyProGlnGlyVally 859
QY      257 GGGCGGTTCAGGCACCTCCAGACTGCTTCTGCTCGCTGCTGCTCCAGAGCTCGCGCTC 198
Db      859 sGlyGluArgGlySerPro-----GlyGlyProGlyAlaGlyAlaGlyPh 873
QY      197 TCCTCTGCTGTCGGCCAACTGCTAGGAATCAGCCAGCGCCCATTTCTGCGAGCCCTT 138
Db      873 eProGlyAla-----ArgGlyProProGlyProProGlyAsnAsn----- 886
QY      137 TGTGCGGGTCCAGCTTCTCAGCCCATGCTCAACACTGCTGCTGCGGCGACCTCAGTG 78
Db      887 -GlyAspArgGlyGluSer-GlyProProGlyValProGlyProProGlyHis---ProG 905
QY      77 GGGACACGCTCATCATCAGATCCTGGCGCA----- 46
Db      905 lyProAlaGlyAsnAsnGlyAlaProGlyLysAlaGlyGluArgGlyPheGlnGlyProL 925
QY      45 -----GGCGCGCGGCTGTCCACCGGAGCCAGCGCGTGCAGGCTGGTTCCC 1
Db      925 euGlyProGlnGlyAlaIleGlySerProGlyAlaSerGlyAlaArgGlyProPro 943

RESULT 7
CAL3_HUMAN
ID   CAL3_HUMAN          STANDARD;          PRT;   1466 AA.
AC   P02461; Q15112;
DT   21-JUL-1986 (Rel. 01, Created)
DI   01-JAN-1990 (Rel. 13, Last sequence update)
DE   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(III) chain precursor.
GN   COL3A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin fibroblast;
RX   MEDLINE=89350838; PubMed=2764886;
RA   Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA   Prockop D.J.;
RT   "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT   chain of human type III procollagen. Differences in protein structure
RT   from type I procollagen and conservation of codon preferences.";
RL   Biochem. J. 260:509-516(1989).
RN   [2]
RP   SEQUENCE OF 149-1225 FROM N.A.
RX   MEDLINE=89386015; PubMed=2780304;

```

```

RA   Janeczko R.A., Ramirez F.;
RT   "Nucleotide and amino acid sequences of the entire human alpha 1
RL   (III) collagen.";
RN   Nucleic Acids Res. 17:6742-6742(1989).
RX   [3]
RP   SEQUENCE OF 168-398.
RX   MEDLINE=77134724; PubMed=557335;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of cyanogen
RT   bromide peptides from the amino-terminal segment of type III collagen
RT   of human liver.";
RL   Biochemistry 16:1158-1164(1977).
RN   [4]
RP   REVISIONS.
RA   Seyer J.M.;
RL   Submitted (DEC-1977) to the PIR data bank.
RN   [5]
RP   SEQUENCE OF 399-727.
RX   MEDLINE=7900343; PubMed=687591;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of five
RT   consecutive CNBr peptides from type III collagen of human liver.";
RL   Biochemistry 17:3404-3411(1978).
RN   [6]
RP   SEQUENCE OF 728-964.
RX   MEDLINE=80198282; PubMed=6246925;
RA   Seyer J.M., Mainardi C., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of alpha 1
RT   (III)-CB5 from type III collagen of human liver.";
RL   Biochemistry 19:1583-1589(1980).
RN   [7]
RP   SEQUENCE OF 950-1466 FROM N.A.
RX   MEDLINE=88189827; PubMed=3357782;
RA   Mankoo B.S., Dalgleish R.;
RT   "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL   Nucleic Acids Res. 16:2337-2337(1988).
RN   [8]
RP   REVISION TO 1184.
RX   MEDLINE=89098346; PubMed=3211760;
RA   Molyneux K., Dalgleish R.;
RT   "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL   Nucleic Acids Res. 16:11833-11833(1988).
RN   [9]
RP   SEQUENCE OF 1065-1466 FROM N.A.
RX   MEDLINE=85087944; PubMed=6096827;
RA   Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA   Rosenbloom J., Myers J.C.;
RT   "Molecular cloning and carboxyl-propeptide analysis of human type III
RT   procollagen.";
RL   Nucleic Acids Res. 12:9383-9394(1984).
RN   [10]
RP   SEQUENCE OF 965-1200.
RX   MEDLINE=81208139; PubMed=7016180;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of alpha
RT   1(III)-CB9 from type III collagen of human liver.";
RL   Biochemistry 20:2621-2627(1981).
RN   [11]
RP   SEQUENCE OF 1176-1466 FROM N.A.
RX   MEDLINE=85157600; PubMed=2579949;
RA   Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
RT   "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT   (III) collagen. Partial characterization of the 3' end region of the
RT   gene.";
RL   J. Biol. Chem. 260:4357-4363(1985).
RN   [12]
RP   SEQUENCE OF 1161-1200 FROM N.A.
RX   MEDLINE=86187804; PubMed=3754462;
RA   Miskulin M., Dalgleish R., Kluge-Beckerman B., Rennard S.I.,
RA   Toistoshkev P., Brantly M., Crystal R.G.;
RT   "Human type III collagen gene expression is coordinately modulated
RT   with the type I collagen genes during fibroblast growth.";
RL   Biochemistry 25:1408-1413(1986).

```

RN [13]
 RP SEQUENCE OF 1-170 FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE=88303360; PubMed=3405773;
 RA Toman D., Ricca G., de Crombrughe B.;
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
 of human prepro alpha 1(III) collagen.";
 RL Nucleic Acids Res. 16:7201-7201(1988).
 RN [14]
 RP SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Well D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 procollagen gene (COL3A1).";
 RL Gene 78:255-265(1989).
 RN [15]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97259599; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I; II, III, and XI), fibril-
 associated collagen (type IX), and network-forming collagen (type X)
 cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [16]
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
 RX MEDLINE=93293988; PubMed=8514866;
 RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinert C.,
 Earley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,
 Cole C.W., Jaakkola P., Rytanen M., Pearce W.H., Yao J.S.T.,
 Rajamanna K., Smulders S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
 Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
 in the triple-helical domain of type III procollagen are an
 infrequent cause of aortic aneurysms.";
 RL J. Clin. Invest. 91:2539-2545(1993).
 RN [17]
 RP VARIANT THR-698.
 RX MEDLINE=91045136; PubMed=2235526;
 RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
 Wu Y., Ganguly A., Prockop D.J.;
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
 RL Nucleic Acids Res. 18:6180-6180(1990).
 RN [18]
 RP VARIANT AORTIC ANEURYSM ARG-786.
 RX MEDLINE=91056145; PubMed=2243125;
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family
 with aortic aneurysms.";
 RL J. Clin. Invest. 86:1465-1473(1990).
 RN [19]
 RP VARIANT EDS-IV ARG-828.
 RX MEDLINE=89109135; PubMed=8411057;
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
 RT "The substitution of glycine 661 by arginine in type III collagen
 produces mutant molecules with different thermal stabilities and
 causes Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 30:690-693(1993).
 RN [20]
 RP VARIANT EDS-IV SER-957.
 RX MEDLINE=94016385; PubMed=2492273;
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
 RT "A single base mutation that substitutes serine for glycine 790 of
 the alpha 1 (III) chain of type III procollagen exposes an arginine
 and causes Ehlers-Danlos syndrome IV.";
 RL J. Biol. Chem. 264:1349-1352(1989).
 RN [21]
 RP VARIANT EDS-IV VAL-960.
 RX MEDLINE=95268429; PubMed=7749417;
 RA Tromp G., de Paeppe A., Nuytinck L., Madhathari S.L., Kuivaniemi H.;
 RT "Substitution of valine for glycine 793 in type III procollagen in
 Ehlers-Danlos syndrome type IV.";
 RL Hum. Mutat. 5:179-181(1995).
 RN [22]

RP VARIANT EDS-IV GLU-1014.
 RX MEDLINE=92316511; PubMed=1352273;
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
 Pope F.M.;
 RT "A single base mutation in the gene for type III collagen (COL3A1)
 converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
 syndrome type IV. An unaffected family member is mosaic for the
 mutation.";
 RL Hum. Genet. 89:414-418(1992).
 RN [23]
 RP VARIANT EDS-IV ASP-1050.
 RX MEDLINE=90037070; PubMed=2808425;
 RA Tromp G., Kuivaniemi H., Stollie C.A., Pope F.M., Prockop D.J.;
 RT "Single base mutation in the type III procollagen gene that converts
 the codon for glycine 883 to aspartate in a mild variant of
 Ehlers-Danlos syndrome IV.";
 RL J. Biol. Chem. 264:19313-19317(1989).
 RN [24]
 RP VARIANT EDS-IV VAL-1077.
 RX MEDLINE=91374480; PubMed=1895316;
 RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,
 Pope F.M.;
 RT "Characterisation of a glycine to valine substitution at amino acid
 position 910 of the triple helical region of type III collagen in a
 patient with Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 28:458-463(1991).
 RN [25]
 RP VARIANT EDS-IV GLU-1173.
 RX MEDLINE=93022543; PubMed=1357232;
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;
 RT "Alignment Scores:
 Pred. No.: 3,87e-09 Length: 1466
 Score: 351.50 Matches: 289
 Percent Similarity: 29.85% Conservative: 57
 Best Local Similarity: 24.94% Mismatches: 391
 Query Match: 5.65% Indels: 424
 DB: 62
 US-09-759-143-110 (1-3410) x CAL3_HUMAN (1-1466)
 Qy 3051 AACAGCGTGTGGAGCTGGTGGGAAAGTTGGGGTAGGGGAAAGTTGGGGTAGGGGA 2992
 Db 161 LysSerGlyValAlaValGlyGlyLeuAlaGlyTyrProGlyProAlaGlyProGly 180
 Qy 2991 AATTTGGGAGTGCCTTCATCAGCCAGTCCCTAGAGAGAGTAGAGGGG----- 2943
 Db 181 -----ProGlyProProGlyProGlyProGlyHisProGly 193
 Qy 2942 -AGTGAAGTGGGGGAACAGGCTGGGCCAAGAGAGAGGGTGGTTAGGGAAGCGTT 2884
 Db 193 rProGlySerProGlyTyrGlnGlyProGlyGluProGlyGluProGlyAlaGly---ProSe 212
 Qy 2883 GAGACCTGAAGCCGCCACCCCTCTACCTTCCTTCAACACCCCTAACCTTGGGTAAACAGCATTT 2824
 Db 212 rGlyProGlyProGly----- 219
 Qy 2823 GGAATTATCATTTGGGATGAGTAGAATTCACAGGTCCTGGTTAGGCATTTTGGGGGC 2764
 Db 220 -----AlaIleGlyProSerGlyProAlaGlyLysAspGlyLysArgProGly 237
 Qy 2763 CAGACCCAGAGAGAAAGATTCTGGCAATGATCAGCCCAATGACAGCATCTCTCAGGGG 2704
 Db 237 yArgProGlyGluArgGlyLeu-----ProGlyProProGlyIleGly 252
 Qy 2703 ACCTGATTGTTGGGATCCCCACCCCTACCCAAATATTAGACACACACAGA----- 2651
 Db 252 yProAla-----GlyIle-ProGlyPheProGlyMetLys---GlyHisArgGlyPheA 269
 Qy 2650 -----AAAGCTAGCAATGGAATCCCTCTTACTTCTTAAATAAATAAG 2608
 Db 269 spGlyArgAsnGlyGluLysGlyGluThrGlyAlaProGly-----L 283

QY	2607	TTAAATATTAAATGCTGTGTCTCTGATGCAACAGAGAACACAGCCACATCC	2548	Db	553	-----ProGly-SerGlnGlyGluSerGlyArgPro-----	563
Db	283	eulysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaProGlyProMetGlyProArgG	303	QY	1529	CTCGTATTGGGCGAGGAACACCTCTCTCCCGTGTAGAGGAGGAGCCAGTGTGTAGG	1470
QY	2547	TGATAAAGGTAGAGGGGGTGGATCAGCAAAAGACAGTGTGTGGCTGAGGGAGCC	2488	Db	564	-----GlyProGlyProSerGlyProArgGlyGlnPro-----	575
Db	303	lyAlaProGlyGluArgGlyArgProGlyAlaAlaGlyAlaArgGlyAsn-	322	QY	1469	GCAGGATCTGACGGCTGAGAAGTGAAACCGGTGAGGGCGGTGAAGCTGTACACCGG	1410
QY	2487	TGTTCTTGTGTGCTTGCCTCAGGACTCTTCCCTACAAATAGTCATATGTC	2428	Db	576	-----GlyValMetGlyPhe-----ProGly	582
Db	322	-----	322	QY	1409	CCACTGTGGCAGAGCATGTGGCACCGGCGAGCCACAGGAAAGCTGCCACACTGGCCA	1350
QY	2427	CCATGGAGAGTGTTCATCTCTAGAACTCCCATGCAAGAGCTACATTAACGAACTGC	2368	Db	583	ProLysGlyAsnAspGly-----AlaProGlyLys-----	592
Db	322	-----	322	QY	1349	AATAGACTGTCGAGTGCAGTCCGATCCACACCGGTCTCATGACAGAGAGACCA	1290
QY	2367	AGTTAAGGGCTTACA-----GATGGGAACACAGTGCATGAGTTATTACGTCCCA	2314	Db	593	-----AsnGlyGluArgGlyProGlyGlyProGlyProGlnGlyProGlnGlyProPro	608
Db	323	-----AspGlyAlaArgGlySerAspGlyGlnProGly-----Prop	335	QY	1289	GGGATGGCGCACTGCAGGA-----ACAGCCCGAGGCTGCCCA	1251
QY	2313	AAACCCCTTCTCTAGGTGTCTTCAACTAGGAGGCTAGCTGTTAACTGAGCCTGGGTA	2254	Db	609	GlyLysAsnGlyGluThrGlyProGlnGlyProGlyProThrGlyProGlyGlyAsp	628
Db	335	roGlyProPro-----GlyThrAlaGlyPhePro-----Glys	346	QY	1250	TCCGAACCCCTTCATCATAGTGTCTCCGGGCTCGGTCCCGGCTCAGCTCTGGCACGC	1191
QY	2253	ATCCACTGCAGAGTCCCGCATTCAGTGCATGGAGCCCT-----TCTGGCCCTCCCTG	2200	Db	629	Lys-----GlyAspThrGlyProGlyProGlyProGlnGlyLeuGlnGlyLeu	643
Db	346	erProGlyAlaGlyGlyGluValGlyProAlaGlySerProGlySerAsnGlyAlaProG	366	QY	1190	CCTGGTACA-----	1182
QY	2199	TATAAGTCCAGACTGAACCCCTTGGAGGCTCCAGTCAGGAGCCCTAGAGACTGGG	2140	Db	644	ProGlyThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLysGly	663
Db	366	ly-----GlnArgGlyGluProGlyProGlnG	375	QY	1181	-----GCCCTCGCCACCA	1167
QY	2139	GAGAG-----AGGAGAGGAGGCCCCAGCCCTGTCAGCTACGACCTCAGCAG	2086	Db	664	AspAlaGlyAlaProGlyAlaProGlyLysGlyAspAlaGlyAlaProGlyGluArg	683
Db	375	lyHisAlaGlyAlaGlnGlyProProGlyProGlyPhe-----AsnGlyS	391	QY	1166	AATCGTGTAAACAGCGCTGAAGTCTAGTGCATCCAGCTCCAGCTCCAGCTCCAGCA	1107
QY	2085	CACAGGTGGCAGCAGAGCCACATTTACTTTGGCAGCAACAGAACTGGCGCCAGCCC	2026	Db	684	GlyPro-----ProGlyLeuAlaGlyAlaProGlyLeu	694
Db	391	erProGlyGlyLysGlyGlu-----MetGlyProAlaG	402	QY	1106	AGACCGCGCGAGGTGGGGCATGGCGGACACAGCTGTGTGACCGGGGAGAGCAGG	1047
QY	2025	GGCAGCCCCATGGGCTAACAGGAGGCGGAGCTGGACCCAGTGCAGCAGGCCCCC	1966	Db	695	ArgGlyGlyAlaGlyProProGlyProGlyLysGlyLysGlyAlaAlaGlyProGly	714
Db	402	lyLeProGlyAlaProGlyLeuMetGlyAlaArgGlyPro-----ProGlyProAlaG	420	QY	1046	CGCCAGGTCC-----	1035
QY	1965	CCCAATGCTGGAAGTTTCTACGCTG-----AGTATTGGCCAGTGGCTCTTGTCAA	1910	Db	715	ProGlyAlaAlaGlyThrProGlyLeuGlnGlyMetProGlyGluArgGlyGlyLeu	734
Db	420	lyAlaAsnGlyAlaProGlyLeuArgGlyGlyAlaGlyGluProGlyLysAsn-----	437	QY	1034	-----GGAAGCCAGCGCGCCCGCATGGCAGCAGT	1002
QY	1909	TACTACCTGTAGCAAGTAATAGCCAGCCCA-----	1871	Db	735	GlySerProGlyProLysGlyAspLysGlyGluProGlyGlyProGlyAlaAspGlyVal	754
Db	438	-----GlyAlaLysGlyGluProGlyProArgGlyGluArgGlyGluAlaG	453	QY	1001	GGCGGACA-----AGAGGGGCGCCACACCCCTTCTGTGCTCGGTGGGCGCCAGC	948
QY	1870	GCTGTGGCAGACACATATAGGCTGACAGCTGGTGTAGCTGCAACAATGGAGCCCAT	1811	Db	755	ProGlyLysAspGlyProArgGlyProThr-----GlyProIle	767
Db	453	lyLeProGlyValProGlyAlaLysGlyLysAspGlyLysAspGlySerProGlyGluP	473	QY	947	CTGCTCTCTCAGCCAGCAGTGTGGCTGTAGCAGGTGAGGAAGATGAGGTGAGCA	888
QY	1810	AAACAGGATGGGCACTGGGACAGGAG-----AGGCATCTATCCAGTGGCA-----	1758	Db	768	GlyProProGlyProAla-----GlyGlnProGlyAspLys	779
Db	473	roGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyAlaProGlyPheArgGlyP	493	QY	887	GGCCAAAGAGCAGTCTCTCTGGTGGCCAGGT-----AGGGGGCGGAGGCGACTGTGTCCC	831
QY	1757	-----GGTCCAGCAGATCCCGCGCGGAGAACCCCTGGCTCGGTGGCTCACCCAC	1703	Db	780	GlyGluGlyGlyAla-----ProGlyLeuProGlyIleAlaGlyPro-----	793
Db	493	roAlaGlyProAsnGlyIleProGlyGlyLysGlyProAlaGlyGluArgGlyAlaProG	513	QY	830	AGTCAATGGCAGGAGGAGTGGCCAGGAGCCCGCCAGCTGATCATGAAGCATAGA	771
QY	1702	CACCACAGTACGAGACATCACAGGAGGCGCCCGCAGAGC-----CGGGTGGAG	1650	Db	794	-----ArgGlySerProGlyGlu-----	802
Db	513	lyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArgAspGlyValProGlyGlyP	533	QY	770	CAGAGTAGGCTGGCGACAGTGT-----CCGGGT-----CCCGAAGAGGTTCAGAGCAGG	717
QY	1649	GTGGAGGAGCCACTGCTCCAGCAGCCCGCTGTTCATTAGGAGGAGGAGCTCCAGCT	1590	Db	803	ThrGlyProProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGluProGly	822
Db	533	roGlyMetArgGlyMetProGlySerProGlyGlyProGlySerAspGlyLysProGly-	552	QY	716	CCTCAGTGGAGTGAAGCACACCTGGCCACAGAGTCCAGCAGCCCGCCAGCATGA	657
QY	1589	TAGGCGCTGGCAGGAGTGTCTCATCAGGCTGTCTCTACTGCTAGCAGCTCCAGTGTCCC	1530				

```

Db      823 GlyLysGlyGluArgGlyAlaProGlyGluLysGlyGlyGlyPro---ProGlyVal 841
QY      656 GCAGTCCAGCTCCAGGG-----GCTGGGATCCGGGCACAGCAGCCTGTAGCC 606
Db      842 AlaGlyProProGlySerGlyProAlaGlyProGlyProGlyProGlyValLysGly 861
QY      605 AGCCGCGCCTGGGATGAGAAAGAGCTCAGCAGGATGCCAAGACAGTGCCTCAGATGA 546
Db      862 -----GluArgGlySerProGlyGlyPro----- 869
QY      545 AGGGCGCGCGCGGCATAGCTCCACGCCAGTGTCTACTGCTGAGCTAGGACGGGA 486
Db      870 -----GlyAlaAlaGlyPhe-----ProGlyAlaArg 878
QY      485 CACAGACAGCGCCAGCAGCTGGACCAATGCCCGCAGCAGCATGCTCACTCTCTCTA 426
Db      879 GlyLeuProGlyPro-----ProGlySerAsnGlyAsnProGly 891
QY      425 CCCCCATTCACAGCAGCAGCGCGGCACATAGTGTATGCTCGCGCCAAACACCTCA 366
Db      892 ProProGlyProSerGlySer-----Pro 899
QY      365 GGCCAAAGTTAGCAGCTTGACCAAGCAGCTGGCTTTCCGGTGGCCGACAGCGCGC 306
Db      900 GlyLys-----AspGlyProProGlyProAlaGlyAsnThrGlyAlaPro-----Gly 915
QY      305 TCACCCACAGCCTCTCGACCATAGTGG-----CCAGCGGGTAGGGCTCAGG 258
Db      916 SerProGlyValSerGlyPro-LysGlyAspAlaGlyGlnProGlyGluLysGlySerPr 935
QY      257 GGCGCCTTCAGGCACCTCAGAACTGCTTCTCGCTCGGCTGTCTCCA----- 212
Db      935 oGlyAlaGlnGlyProPro-----GlyAlaProGlyProLeuGlyIle 949
QY      211 -----GAAGCTCGCGCTCTCTCTCTCTCTCTCGCGCAACTGCCTAG 171
Db      949 eAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetProGly---ProAr 968
QY      170 GAATCAGCAGCGCGCCATTTCTGCCAGCCCTTTGGTGGCGGTCCAGCTTCTCAG----- 116
Db      968 gGlySerProGlyProGlnGlyValLysGlyGluSerGlyLysProGlyAlaAsnGlyLe 988
QY      115 -----CCATGCTCAACACTGCTGCTGGCGCACCTCA 81
Db      988 uSerGlyGluArgGlyProProGlyProGln-GlyLeuProGlyLeuAlaGly---ThrA 1007
QY      80 GTGGGACAGCTCTCATCACTCAGATCCTGGC----- 49
Db      1007 laGlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeuProGlyArgAspGlyS. 1027

QY      48 -----CGAGCGCGCGGCTGTACCCCGGAGCC 22
Db      1027 erProGlyGlyLysGlyAspArgGlyGluAsnGlySerProGlyAla 1042

RESULT 8
CALL CANFA
ID      CALL_CANFA      STANDARD;      PRT; 1460 AA.
AC      Q9XJ7;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Collagen alpha 1(I) chain precursor.
GN      COL1A1.
OS      Canis familiaris (Dog).
OC      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX      NCBI_TaxID=9615;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      Tissue-Skin;
RA      Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
RT      "Sequence of normal canine COL1A1 cDNA.";
RL      Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

```

```

CC      -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC      (FIBRILLAR FORMING COLLAGEN).
CC      -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC      -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC      UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC      -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; AF153062; AAD34619.1; -.
DR      InterPro; IPR000087; Collagen.
DR      InterPro; IPR000885; Fib_collagen_C.
DR      Pfam; PF01391; Collagen; 18.
DR      Pfam; PF01410; COLFI; 1.
DR      ProDom; PD000007; Collagen; 1.
DR      ProDom; PD002078; Fib_collagen_C; 1.
DR      SMART; SM00038; COLFI; 1.
DR      SMART; SM00214; VWFC; 1.
DR      PROSITE; PS01208; VWFC; 1.
KW      Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW      Glycoprotein; Collagen; Signal.
FT      SIGNAL 1 22
FT      PROPEP 23 157
FT      CHAIN 158 1214
FT      PROPEP 1215 1460
FT      DOMAIN 34 92
FT      DOMAIN 158 174
FT      DOMAIN 175 1188
FT      DOMAIN 1189 1214
FT      SITE 741 743
FT      SITE 1089 1091
FT      CARBOHYD 1361 1361
SQ      SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Alignment Scores:
Pred. No.: 4,51e-09 Length: 1460
Score: 350.00 Matches: 277
Percent Similarity: 31.11% Conservativeness: 50
Best Local Similarity: 26.36% Mismatches: 365
Query Match: 5.62% Indels: 360
DB: 1 Gaps: 56

US-09-759-143-110 (1-3410) x CALL_CANFA (1-1460)

QY      2591 CTGTGCTCTCTGTGAT---GGCAACAGAGGACCAAGCCACATCTGTGATAAAGGTAA 2535
Db      56 IleCysValCysAspAsnGlyAsnValLeuLeuCysAspValIleCysAspGlu----- 73
QY      2534 GAGGGGGGTGATCATCAGCAAAAAGACAGCTGCTGGCTGAGGGACCTGTTCTGTGTG 2475
Db      74 -----ThrLysAsnCysProGlyAlaGlnValProProGly-GluCy 87
QY      2474 TTGCCCCCTCAGGACTCTTCCCTACAAATAAGTCATATGTTCAATCCCATGGAGAGTG 2415
Db      87 SCYsPro-----ValCys----- 91
QY      2414 TTTCATCTAGAAACTCCCATGCAAGAGCTACATTAAACAGCTGACAGTTAAGGGCT 2355
Db      92 -----ProAspGlyGluAlaSerProThrAspGlnGluThrGlyVa 106
QY      2354 T----- 2354
Db      106 IGLuGlyProLysGlyAspThrGlyProArgGlyProArgGlyProAlaGlyProProGly 126
QY      2353 -AGAGATGGGAACCCAGGTGACTGAGTTTATTCAGCTCCCAAAACCCTTCTCTAGGTGT 2295

```

Db	126	YArgAspGlyIleProGlyGlnProGlyLeuProGlyProProGlyProPro	143
QY	2294	GTCTCAACTAGGAGCTAGCTGTAAACCTGAGCCTGGTAAATCCACTGCAGATGCC	2235
Db	144	-----GlyProProGlyProProGly	150
QY	2234	GCATTCCAGTCATGGAGCCCTCTCT-----GGCCTCCCTGTATATAGTCCAGACTGA	2184
Db	150	YLeuGlyGlyAsnPheAlaProGlnMetSerTyrGly-----TyrAspGly	165
QY	2183	AAACCCCTTGAAGCCCTCCAGTCAGGCAGCCCTAGAGACTGGGGAGAGAGAGGAC	2124
Db	165	uLysSerThrGlyGlyIleSerValProGlyProMetGlyProSerGlyProArgGlyLe	185
QY	2123	GCCCGAGCCCGAGCTGTGACCTACGCACCTCAGCAGCAGGAGTGGCAGCAGAGACC	2064
Db	185	uProGlyProProGlyAlaPro-----GlyProGlnGlyPheGlnGlyProProGlyGluPr	204
QY	2063	ACATTACTTTGGCAGCAACAGAAATGGCGGCCAGCCCGCCAGCCCTAGGCTAAACAG	2004
Db	204	oGlyGluProGlyAlaSerGlyProMetGlyProArgGlyProProGlyProProGlyLy	224
QY	2003	CAGCGGGAGCTGGGA-----CCAGTGGAGCAGCCCTCCA--	1967
Db	224	sAsnGlyAspAspGlyGluAlaGlyLysProGlyArgProGlyGluArgGlyProProGl	244
QY	1966	-----CCCCAATGTCTGGAAGTTTCTACGCTCAGTATTGG	1929
Db	244	yProGlnGlyAlaArgGlyLeuProGlyThrAlaGly-----LeuProGl	259
QY	1928	CCAAGTCGCTCTTCTCAATATACTCTGTGTAGCAAAAGTAAATGGCAGCAGCCAGGC	1869
Db	259	yMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAspAlaGlyProAl	279
QY	1868	CTGGGGCAGACCATATAGCAGTGCAGTGCAGTGGTGCAGTGCAGTGCAGTGCAGTGC	1809
Db	279	aglyProLysGlyGluProGlySerProGly-----GluAsnGlyAlaProGl	295
QY	1808	ACAGGGAGGGGCA-----CCTGGGACAGCAGGAGCAGTATCCAGGATGGCGA	1758
Db	295	yGln--MetGlyProArgGlyLeuProGlyGluArgGlyArg-----ProGlyAlaPro	312
QY	1757	GGTCAGGACAGATGCCCGCCGCGAA-----CCACCTCGCCCT	1719
Db	313	GlyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAlaAlaGlyProProGlyPro	332
QY	1718	CGTGGGCTCACCACACCACCA-----CGTACGAGACATCA	1681
Db	333	ThrGlyProAlaGlyProProGlyPheProGlyAlaValGlyAlaLysGlyGluAlaGly	352
QY	1680	CAGGCAGAGCCCGCAGCAGCGGGTGGAGTGGAGCAGG-----CCACTGCCCTCCA	1627
Db	353	ProGlnGlyAlaArgLysGlyGluGlyProGlnGlyValArgGlyGluProGlyProPro	372
QY	1626	GCACCCAGCTGCCATTAGGG---AAGGGAGCTCCAGCTTAGGG-----CCTGGCAGG	1576
Db	373	GlyProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnProGlyAla	392
QY	1575	AGCTGTGTCATCAGCTCTCTCTACTGCTAGCAGCTCAGTGTCCCTCGTATTGGGC	1516
Db	393	Lys-GlyAlaAsnGlyAlaPro-----GlyIle--Al	402
QY	1515	AGGAACACTCTCTCTCCGCTGTAGAGGAGGCCAGTGTGTAGGCGCAGATCTGCAGG	1456
Db	402	aglyAlaProGlyPheProGlyAlaArgGly-----ProSerGlyPro-GlnG	418
QY	1455	GCTGAGAAGGTGAACCCGGTGGAGCGGCTGAAGCTGCACCGGCCACACTCTGGGAC	1396
Db	418	lyProSerGlyProProGlyProLysGlyAsnSerGlyGluProGly-----AlaProGlyA	437
QY	1395	AGGCATGTGGCAGCCGAGCCAGGAGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGA	1336
Db	437	snLysGlyAspThrGlyAlaLysGlyGluProGlyProThrGly-----	451
QY	1335	GTCCGAATCGCTGCACACCGCGGTCCATGACACAGAGAGAGAGGAGATGGCGCAC	1276
Db	452	-----IleGlnGlyPro--ProGlyProAlaGlyGluGluGlyLysArgGlyAl	467
QY	1275	TCAGAGAACA-----GCCCCAGGCTGCCATCCGAACGC--	1242
Db	467	aArgGlyGluProGlyProThrGlyLeuProGlyProProGlyGluArgGlyGlyProGl	487
QY	1241	-----CTTCATCATAGTCTCCGGGCTCGGTGCGCGGCTCAGCTCTG	1198
Db	487	ySerArgGlyPheProGlyAlaAspGlyValAlaGlyProLysGlyProAlaGlyGluAr	507
QY	1197	GGCAGCCCTGTGTACAGCCCTCGCCACGAATCGTGTAAACACGCGTGAAGTCA	1138
Db	507	gGlySerProGlyProAlaGlyProLysGlySerPro-----	519
QY	1137	AGTGCCATCCAGTGCACAGCTCAGCCACAGAGCCGCGCAGGTGCGGGGCATGCGG	1078
Db	519	-----	519
QY	1077	CACACAGCTGTGTGACAGCCGGGGAACAGCGGCGCCCAAGT-----TCCGGAAA	1030
Db	520	-GlyGluAlaGlyArgProGlyGluAlaGlyLeuProGlyAlaLysGlyLeuThrGlySe	539
QY	1029	GCCAAGCGGGCGGCGATGACAGCAGTGGGGCGCAAGAGGGGGCGCACGCCCTTCT	970
Db	539	rProGlySerProGlyProAspGly-----LysThrGlyProPro--	552
QY	969	GCTGGCTCGGGGGCGCCAGCG-----CTGCCCTCTCAGCCACCA--	930
Db	553	-----GlyProAlaGlyGlnAspGlyArgProGlyProProGlyProProGl	568
QY	929	-----GCAGTGTGGCTGCT	916
Db	568	yAlaArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAla--	585
QY	915	ACCGAGGTGAGGAGATGAGGTGAGCGCCCAAGAGGCACT-----CCTCTGGGTG	862
Db	586	-----GlyGluProGlyLysAlaGlyGluArgGlyValProGlyProProGlyAl	602
QY	861	CCAGGT-----AGGGG	850
Db	602	aValGlyProAlaGlyLysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAlaGl	622
QY	849	CCAGGCACTGTGTCCAGTCAATGGCAGGACAG-----	813
Db	622	yProAlaGlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPheGlnGlyLeuPr	642
QY	812	-GTCAGCCAGGAGCCCGCCCAAGACTGATCATGAGGCATAGACAGTAGGCTGGCGA	754
Db	642	oGlyProAlaGlyProProGlyGluAlaGlyLysProGlyGluGlnGlyValProGlyAs	662
QY	753	CAGTGTGT---CCGGGTCCC-----GGAAGAGCTCAGACAGCAGGCTCAGTGGAGTG	703
Db	662	pLeuGlyAlaProGlyProSerGlyAlaArgGlyGluArg---GlyPheProGlyGluAr	681
QY	702	AACACACCTGGCCACAGAGTCCAGCAGCCCGCCAGGATGAGCAGTGCAC-----	648
Db	681	gGlyValGlnGlyProProGlyProAlaGlyProArgGlyAlaAsnGlyAlaProGlyAs	701
QY	647	-----GCTCCAGGCGCTGGGATCCCGGACACAGCCCT	613
Db	701	nAspGlyAlaLysGlyAspAlaGlyAlaProGlyAla-----ProGlySerGlnGlyAl	719
QY	612	GCTAGCAGCCCGCCCTTGGGATGAGAAAGAGCTCAGCAGGATGCCCAAGGACAGTGCC	553
Db	719	aProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGly-----Pr	737
QY	552	CAGATGAGGCGCGGGCGCCGATACGCTCCAGCGGCTGCTGCTGCTGCTGCTGCTGCTG	493
Db	737	oLysGlyAspArgGlyAspAlaGlyProLysGlyAlaAspGly-----SerProGl	754

QY 492 AGCG---GGACACAGA-----CCAGGCCCGACACTGCACCA 460
 Db 754 YLysAspGlyValArgGlyLeuThrGlyProIleGlyProGlyProAla-----GI 772
 QY 459 ATGCCCA----- 453
 Db 772 YAlaProGlyAspGlyGluAlaGlyProSerGlyProAlaGlyProThrGlyAlaAr 792
 QY 452 ----GCACCATGTGATGAATCTCTCTACCCCACTTCCAGCAGCAGCGCGCAC 397
 Db 792 gGlyAlaProGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGly-- 811
 QY 396 TAGGTGATGCTCGGCCCAACACACTCCAGGC-----CAAGGTTAGCAGG 349
 Db 812 -----ProGlyGlyAlaAspGlyGlnProGlyAlaLysGI 823
 QY 348 TTGACCAGCAGAGCTGGGCTTTCCGGTGCAGCAGCGCGCTCACCCACAGCCTCTGG 289
 Db 823 YGluProGlyAspAlaGlyAlaLysGlyAspAla-----GlyProGlyProAlaGI 841
 QY 288 ACCATAGTGGG---CCAGCGGGTAGGCTCAGCGGGCGGCTTCAGGCACCTCCAGAACTGC 232
 Db 841 YPro-ThrGlyProProGlyProIleGlyAsnValGlyAlaProGly---ProLysGlyA 860
 QY 231 TTGCTCTCGGCTCTGCTCCAGAAAGCTCGCGCCTCTCTCTGCTCGCGCCCAACTGCCTA 172
 Db 860 laArgGlySerAlaGlyProProGlyAlaThrGlyPhePro-GlyAlaAlaGlyArgVal 879
 QY 171 GGAATCAGCCAGCGGCCCATTTCTGCACGCCCTTGGTCCGCGCTCCAGCTTCTCAGCCCA 112
 Db 880 Gly-----ProGlyProSerGlyAsnAlaGlyPro 890
 QY 111 TGCTCAACACTCTGCTGTGGGCAC-----CTCAGTGGGACACAGCTCT 67
 Db 891 ProGlyProProGlyProAlaGlyLysGluGlyLysGlyAlaArgGlyGluThrGly 910
 QY 66 CATCACTCAGATCTGCGCGCA-----GCAGCGCGG 37
 Db 911 ProAlaGlyArgProGlyGluValGlyProProGlyProProGlyProAlaGlyGluLys 930
 QY 36 CTGTCACCGGAGCC 22
 Db 931 GlySerProGlyAla 935

RESULT 9
 CA44_HUMAN
 ID CA44_HUMAN STANDARD; PRT: 1690 AA.
 AC P53420;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 4(IV) chain precursor.
 GN COL4A4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=95014445; PubMed=7523402;
 RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Readers S.T.;
 RT "Complete primary structure of the human type IV collagen alpha 4(IV)
 RT chain. Comparison with structure and expression of the other alpha
 RT (IV) chains.";
 RL J. Biol. Chem. 269:26172-26177(1994).
 RN [2]
 RN SEQUENCE OF 1-23 FROM N.A.
 RX MEDLINE=98196854; PubMed=9537506;
 RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
 RA Ninomiya Y.;
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
 RT alpha4(IV) collagen chains are arranged head-to-head on chromosome

RT 2q36.";
 RL FEBS Lett. 424:11-16(1998).
 RN [3]
 RP SEQUENCE OF 1219-1690 FROM N.A.
 RC TISSUE=Eye;
 RX MEDLINE=93374047; PubMed=8365481;
 RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;
 RT "cDNA isolation and partial gene structure of the human alpha 4(IV)
 RT collagen chain.";
 RL FEBS Lett. 330:122-128(1993).
 RN [4]
 RP SEQUENCE OF 1407-1507 FROM N.A.
 RX MEDLINE=93054733; PubMed=1429714;
 RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
 RT alpha 4 chain of basement membrane collagen type IV and assignment of
 RT the gene to the distal long arm of human chromosome 2.";
 RL J. Biol. Chem. 267:23753-23758(1992).
 RN [5]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97338662; PubMed=9195222;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 RT "The clinical spectrum of type IV collagen mutations.";
 RL Hum. Mutat. 9:477-499(1997).
 RN [6]
 RP VARIANT AS SER-1201.
 RX MEDLINE=95078927; PubMed=7987396;
 RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,
 RA Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,
 RA Smeets H.J.M., Readers S.T.;
 RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)
 RT collagen genes in autosomal recessive Alport syndrome.";
 RL Nat. Genet. 8:77-82(1994).
 RN [7]
 RP VARIANT BFH GLU-897.
 RX MEDLINE=96379660; PubMed=8787673;
 RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,
 RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;
 RT "Benign familial hematuria due to mutation of the type IV collagen
 RT alpha4 gene.";
 RL J. Clin. Invest. 98:1114-1118(1996).
 RN [8]
 RP VARIANTS AS, AND VARIANTS.
 RX MEDLINE=99011253; PubMed=9792860;
 RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,
 RA Cochat P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;
 RT "Determination of the genomic structure of the COL4A4 gene and of
 RT novel mutations causing autosomal recessive Alport syndrome.";
 RL Am. J. Hum. Genet. 63:1329-1340(1998).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
 CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
 CC COCHLEA, LUNG AND BRAIN.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF TYPE II AUTOSOMAL
 CC RECESSIVE FORM OF ALPORT SYNDROME (AS), AN HEREDITARY

QY 2585 CTCTGTGATGGCAACAGAGACCAACAGGCCAC-----ATCCTGATAAAGGTAA 2535
Db : : : : :
QY 401 aGlyMetIleGlyProProGlyProGlnGlyPheProGlyLeuProGlyGly 421
Db : : : : :
QY 2534 GAGGGGGTGGATCAGCAAAAGACAGTGTCTGGGCTGAGGGACCTGTTCTGTGTG 2475
Db : : : : :
QY 421 uAlaGlyLeuProGlyArgProAspSerAlaProGly----- 433
QY 2474 TTGCCCCCTCAGGACTCTTCCCTTACAATAAAGTCATATGTTCAATCCCATGAGGAGTG 2415
Db : : : : :
QY 433 ----- 433
QY 2414 TTTTCATCTAGAAATCCCATGCAAGAGCTACATTAAACGAAGCTGAGGTTAAGGGCT 2355
Db : : : : :
QY 433 ----- 433
QY 2354 TAGAGATGGAAACAGGTGACTGAGTTATTACAGCTCCCAAAACCCCTTCTCTA---GG 2298
Db : : : : :
QY 434 -LysProGlyLysProGly-----SerProGlyLeuProGlyAlaProGly 448
QY 2297 TGTGCTCAACTA---GGAGCTAGCTGTAAACCTGAGCCTGGGTAAATCCACCTGCAGA 2241
Db : : : : :
QY 448 yLeuGlnGlyLeuProGlySerValIleTyrCysSerValGlyAsnProGlyProGly 468
QY 2240 GTCCCCGCATTCAGTGCATGGAGCCCTTCTGCCTCCCTGTATAGTCCAGACTCAAAAC 2181
Db : : : : :
QY 468 nGlyIleGlyLysValGlyProProGlyGlyArgGly-----ProLysGlyGly 486
QY 2180 CCCTCTGGAAAGCCTCAGTCAGGACCCCTAGAGACTGGGGAGAGAGAGGAGCGCC 2121
Db : : : : :
QY 486 sGlyAsnGlyLeuLysCysAlaCysGluProGlyPro-----MetGlyProPr 502
QY 2120 CCAGCCCCAGCTGTGAGCTAGCAGCTCAGCAGTGGGGTGGCAGCAGAGCCACA 2061
Db : : : : :
QY 502 oGlyProProGlyLeuProGlyArg---GlnGlySerLysGlyAspLeuGlyLeuProGly 521
QY 2060 TTACTTTGGAGCAACAAAGAACTGGCGGCA-----GCCCGGAGCGCC----- 2018
Db : : : : :
QY 521 yTrpLeuGlyThryLysAspProGlyProProGlyAlaGluGlyProProGlyLeuPr 541
QY 2017 -----CATGGGGCTAAC----- 2006
Db : : : : :
QY 541 oGlyLysHisGlyAlaSerGlyProProGlyAsnLysGlyAlaLysGlyAspMetValVa 561
QY 2005 -----AGGAGCGGGAGCTGGACCCAGTGGAGCGGCCCTCCA----- 1967
Db : : : : :
QY 561 lSerArgValLysGlyHisLysGlyGluArgGlyProAsp-----GlyProProGlyPh 579
QY 1966 -CCCAATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAAGTGGCTCTTGTCAATA 1908
Db : : : : :
QY 579 eProGlyGlnProGlySerHisGlyArg-----AspGlyHisAlaGly----- 593
QY 1907 CTACCTGTGTAGCAAAATAATGCGACCA----- 1876
Db : : : : :
QY 594 -----GluLysGlyAspProGlyProProGlyAspHisGluAspAla 608
QY 1875 CCCAGGCTGGCAGACACATATAGCAGTGCAGACTGGCTGAGCTGGACATGAG 1816
Db : : : : :
QY 608 rProGlyGlyLysPheProGlyProLeuGlyProProGlyLysAlaGlyProValGly 628
QY 1815 CCCAAACAGGATGGGGCCACTTGGGACAGCAGGAGGAC----- 1773
Db : : : : :
QY 628 yPro---ProGlyLeuGlyPheProGlyProProGlyGluArgGlyHisProGlyValPr 647
QY 1772 -----TATCCAGGATGGGAGTCCAGCAGATGCCCGGGCCGAGAA----- 1731
Db : : : : :
QY 647 oGlyHisProGlyValArgGlyProAspGlyLeuLysGlyGlnLysGlyAspThrIleSe 667
QY 1730 -----CCACCTGGCC----- 1720
Db : : : : :
QY 667 rCysAsnValThrTyrProGlyArgHisGlyProProGlyPheAspGlyProProGlyPr 687
QY 1719 TCGGTG-----GG 1712

Db : : : : :
QY 687 oLysGlyPheProGlyProGlnGlyAlaProGlyLeuSerGlySerAspGlyHisLysGly 707
QY 1711 CTCACCCACCACACACCTAGGAGACATCACAGCAGAGCCCGGAGAGCGGGGT-- 1654
Db : : : : :
QY 707 yArgProGlyThrProGlyThr-----AlaGluIleProGlyProProGlyPh 723
QY 1653 -----GGAGGT-----GGAGCAGGCCACTG---CC 1631
Db : : : : :
QY 723 eArgGlyAspMetGlyAspProGlyPheGlyGlyLysSerSerProValGlyPr 743
QY 1630 TCCACAGCCACCGTCTCCATTAGG-----AAGGAGCTCCAGC----- 1591
Db : : : : :
QY 743 oProGlyProProGlySerProGlyValAsnGlyLysGlyIleProGlyAspProAl 763
QY 1590 -----TTAGGG-----CCTGGCAGAGAGCT 1571
Db : : : : :
QY 763 aPheGlyHisLeuGlyProProGlyLysArgGlyLeuSerGlyValProGlyIleLys-G 783
QY 1570 GGTCACTAGGCTGCTCACTCACTAGCAGCTCCAGTGTCCCTCGGTATT----- 1521
Db : : : : :
QY 783 yProArgGlyAspProGlyCysProGlyAlaGluGlyProAlaGlyIleProGlyPheL 803
QY 1521 ----- 1521
QY 803 euGlyLeuLysGlyProLysGlyArgGluGlyHisAlaGlyPheProGlyValProGlyP 823
QY 1521 ----- 1521
QY 823 roProGlyHisSerCysGluArgGlyAlaProGlyIleProGlyGlnProGlyLeuProG 843
QY 1520 ---TGGGAGGAAACACTGCTCTCCGCTGTGTAGAGGAGGCCA-----GTGTGTAGGG 1469
Db : : : : :
QY 843 yTyrProGlySerProGlyAlaProGlyGlyLysGlyGlnProGlyAspValGlyProP 863
QY 1468 CAGGATCTCAGGCTGAGAAGGTGAACCCGCTGAGGCGGCTGAAGCTGTCAACACGCC 1409
Db : : : : :
QY 863 roGlyProAlaGlyMetLys-----GlyLeuProGlyLeuProGlyArgP 878
QY 1408 CACACTGTGGGACAGCATGTGGCAGCCAGCCAGGAAAGCTGCCACACTGGCCAA 1349
Db : : : : :
QY 878 roGlyAla-----HisGlyProPro-----GlyLeuProGlyIleProG 891
QY 1348 ATAGACTGCTCAGTGCAGTCCGCAATCGCTGCACAGCCGCTCCA---TGACAGAGAGAAGAC 1292
Db : : : : :
QY 891 yProPheGlyAspAspGlyLeuProGlyProProGlyProLysGlyProArgGlyLeuP 911
QY 1291 CAGGAGATGGCGCACTCAGGAAACAGCCCGCAGGCTGCCATCCCAAGCCCTTCATCATA 1232
Db : : : : :
QY 911 roGly-----PheProGlyPheProGlyGluArgGlyLysProG 924
QY 1231 GTGTCTCCGGGCTCGGTCCCGGCTCAGCTGTGGCAGCCCTGTGTACAGCCCTCCGCC 1172
QY 924 yAlaGluGly-----CysProGlyAlaLys-----GlyGluProGlyGluLysGlyMetS 941
QY 1171 CACGAATCCGTGTAAACAGCAGCTGAAAGTCATGAGTCCATCCAGCTGCACAGCTCAGC 1112
Db : : : : :
QY 941 eGlyLeuPro----- 946
QY 1111 CACGAAGAGCCGGCAGGGTCCGGGGCATGCGGC----- 1077
Db : : : : :
QY 946 sArgGlyLeuArgGlyAlaLysGlyAlaIleGlyProProGlyAspGluGlyGluMetA 966
QY 1076 -----AGCACAGTGGTCCAGCCGGGAGAGAGCGGCCCGCCAGGTCC----- 1035
Db : : : : :
QY 966 lalleSerGlnLysGlyThrProGlyGluProGlyProProGlyAspGlyPheP 986
QY 1034 -----GAAAGCCCAAGCGCGCGGCTGACAGCAGTGGGCGCACAGGAGGG 986
Db : : : : :
QY 986 roGlyGluArgGlyAspLysGlyThrProGlyMetGlnGly-----ArgArg 1002
QY 985 GCCCGACAGCCCTTCTGCTGGCTCGGTGG----- 957
Db : : : : :
QY 957 ----- 957


```
Db 1002 lYglu-----LeuGlyArgTyrGlyProGlyPheHisArgGlyGluProG 1018
QY 956 -----GGCCAGCGTGCCTCTCTCAGCCACAGCAGTGTGGCTGTAC 914
Db 1018 lYgluLysGlyGlnProGlyProGlyProGlyProGlyProGlySerThrGlyLeuA 1038
QY 913 GCAGG---TGAGGAAGATGAGGTGAGCAGGCCAAAGAGGACCTCCCT-----CCTGGGT 863
Db 1038 rgGlyPheIleGlyPheProGlyLeuProGlyAspGlnGlyGluProGlySerProGlyP 1058
QY 862 GCCAGGT-----AGGGGGCCA-----846
Db 1058 roProGlyPheSerGlyIleAspGlyAlaArgGlyProLysGlyAsnLysGlyAspProA 1078
QY 845 --GGCAGCTGG-----TGTCGCCAGTC 827
Db 1078 laSerHisPheGlyProGlyProLysGlyGluProGlySerProGlyCysProGlyH 1098
QY 826 AATGGCAGGAGGAGGTAGCCAGGCAGCCAGCCAGAGCTGATCATGATGAGGACATAGACAGA 767
Db 1098 lsPheGlyAlaSerGlyGlnGlyLeuProGlyIleGlnGlyProArgGlySerProG 1118
QY 766 GTAGGCTGGCAGCAGTGGT-----CCGGGT-----C 740
Db 1118 lYArgProGlyProGlySerSerGlyProGlyProGlyCysProGlyAspHisGlyMetP 1138
QY 739 CCGGA---AGAGGTCAGAGCAGGCGCTCCAGTGGAGTGAAGCAGCACACCTGGCCACAGA- 684
Db 1138 roGlyLeuArgGlyGln-----ProGlyLeuMetGlyAspProGlyProArgG 1154
QY 683 -----AGTCCAGCAGCCCA---CGCCAGGATGAGCAGTGGCCA-----648
Db 1154 lYLeuGlnGlyAspProGlyIleProGlyProGlyIleLysGlyProSerGlySerP 1174
QY 647 -----GCTCCAGGGG 638
Db 1174 roGlyLeuAsnGlyLeuHisGlyLeuLysGlyGlnLysGlyThrLysGlyAlaSerGlyL 1194
QY 637 CCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCGCCCTGGGATGAGAAGAGGCT 578
Db 1194 euHisAspValGlyProGlyProValGlyIleProGlyLeuLysGlyGluArgGlyA 1214
QY 577 CAGCAGATCCCAAGACAGT-----CCAGATGAAGCGCGCGCGGCCATA 527
Db 1214 spProGlySerProGlyIleSerProGlyProArgGlyLysGlyProGlyP 1234
QY 526 GCGTCCAGCGCAGTGGTCACTGGTGAAGCTAGAGCGGACACAGA-----480
Db 1234 roProGlySerGlyProGlyProGlyProAlaGlyAlaThrGlyArgAlaProLysAspI 1254
QY 479 -----CCAGGCCAGCAGTGGACCAATGCCCA-----453
Db 1254 leProAspProGlyProGlyAspGlnGlyProGlyProAspGlyProArgGlyA 1274
QY 452 -----GCACCATGTGTCAT 440
Db 1274 laProGlyProProGlyLeuProGlySerValAspLeuLeuArgGlyGluProGlyAspC 1294
QY 439 GAATCTCTCTTACCCCACTTCCAGCAGCAGCGGCGGCATAGTGTGCTCGCGC 380
Db 1294 ysGlyLeuProGlyProGlyProGlyProGly-----P 1305
QY 379 CAACACACCTCCAGGCCAAAG-----TTAGCAGGTGTGACACAG 341
Db 1305 roProGlyProGlyTyrLysGlyPheProGlyCysAspGlyLysAspGly-GlnLys 1324
QY 340 CAGAGCTGGCTTCGGTGCCGAG---CAGCGGCTCACCACAGCCTCTGGACCAT 284
Db 1325 GlyProMetGlyPheProGlyProGlnGlyProHisGlyPheProGlyProGlyGlu 1344
QY 283 AGTGGG---CCAGCGGGTAGGCTCAGGGGGCGGTTCAGGCACTCCAACTGCTTCGT 227
Db 1345 LysGlyLeuProGlyProGlyArgLysGlyProThrGlyLeuPro-----1360
```

```
QY 226 CTGGCTGCTCTCCAGAGCTGGCGCTCTCTCTCTTC-----TGCCGCCAA 179
Db 1361 -----GlyProArgGlyGluProGlyProProAlaAspValAspCysProArg 1377
QY 178 CTCCTTAGGAATCAGCCAGCGGCCCATTTCTGCCACCCCTTTGGTCCCGGTCCAGCTTCT 119
Db 1378 IlePro-GlyLeuProGlyAlaProGlyMetArgGlyPro-----1390
QY 118 CAGCCCATGCTCAACACCTGCTGCTGGGCGCACCTCAGTGGGACACGCTCTCATCCTC 59
Db 1391 -----GluGlyAlaMetGly---LeuProGly-----MetArgG 1401
QY 58 AGATCTGGCGGAGGCGCGCTGTACCCGAGCAGCGCGTGCAGG 10
Db 1401 yProProGlyProGlyCysLysGlyGluProGlyLeuAspGlyArg 1417

RESULT 10
CAL3_BOVIN STANDARD; PRT: 1049 AA.
ID CAL3_BOVIN
AC P04258;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(III) chain.
GN COL3A1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 1-242.
RX MEDLINE=80026026; PubMed=488906;
RA Fietzek P.P., Allmann H., Rauterberg J., Henkel W., Wachter E.,
RA Kuhn K.;
RT "The covalent structure of calf skin type III collagen. I. The amino
RT acid sequence of the amino terminal region of the alpha 1(III) chain
RT (positions 1-222).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
RN [2]
RP SEQUENCE OF 243-422.
RX MEDLINE=80026027; PubMed=488907;
RA Dewes H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. II. The amino
RT acid sequence of the cyanogen bromide peptide alpha 1(III)CBI,8,10,2
RT (positions 223-402).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
RN [3]
RP SEQUENCE OF 423-571.
RX MEDLINE=80026028; PubMed=488908;
RA Bentz H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. III. The
RT amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
RT (positions 403-551).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
RN [4]
RP SEQUENCE OF 572-808.
RX MEDLINE=80026029; PubMed=488909;
RA Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. IV. The amino
RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
RT (positions 552-788).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
RN [5]
RP SEQUENCE OF 809-947.
RX MEDLINE=80026030; PubMed=488910;
RA Dewes H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. V. The amino
RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
RT (position 789-927).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
RN [6]
```


Db 182 Pro-----GlyAlaProGlyProGlnGlyPheGln----- 191
QY 2092 TCAGCAGCACAGGGTGGCAGACAGCCACATTACTTTGGCAGCACAGAACTGGCGG 2033
Db 192 --GlyProProGlyGluProGlyGluProGlyAlaSerGlyProMetGlyProArgGly 210
QY 2032 CCAGCCCGGAGCCATGGGCTTAACAGGAGGGGAGCTGGGA-----CCC 1985
Db 211 ProAlaGlyProProGlyLysAsnGlyAspGlyGluAlaGlyLysProGlyArgPro 230
QY 1984 AGTCAGGACGCCCTCCA-----CCCCAATGTGTGGAAG- 1950
Db 231 GlyGlnArgGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 250
QY 1949 -----TTTCTACGCTGAGTATTTGGCCAACTGCTCTGT 1914
Db 251 ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAsp----- 263
QY 1913 CAATATACCTGTGTAGCAAGTAAAGCCAGCACCA-----GGCCTGGCGGACACA 1857
Db 264 -----GlyAlaLysGlyGlnProGlyProAlaGlyProLysGlyGlu 277
QY 1856 CCATATAGCAGTCACAGACTGGCTGAGCTGGACAATGGAGCCCAATAACAGGATGGGG 1797
Db 278 ProGlySerProGlyGluAsnGlyAlaProGlyGlnMetGlyPro-----ArgGly 294
QY 1796 CCACCTGGGACAGCAGGAGCACTATCCA-----GGATGGCAGGTCCAGGC 1749
Db 295 LeuProGlyGluArgGlyArgProGlyProSerGlyProAlaGlyAlaArgGlyAsnAsp 314
QY 1748 AGATGCCCCGGCCGGAA-----CCACCTGGCTCGTGGTGGCTCACCCACACACA----- 1696
Db 315 GlyAlaProGlyAlaAlaGlyProProGlyProThrGlyProAlaGlyProProGlyPhe 334
QY 1695 -----CCTACGGAGACATCACAGCAGCAGAGGCCCGCCGAGCGCGGGT 1654
Db 335 ProGlyAlaAlaGlyAlaLysGlyGluThrGlyProGlnGlyAlaArgGlySerGlyGly 354
QY 1653 GGAGGTGGGAGCAGG-----CCACTGCTCCACAGCACCCAGCTCCATTAGG-----AAG 1603
Db 355 ProGlnGlySerArgGlyGluProGlyProProGlyProAlaGlyAlaAlaGlyProAla 374
QY 1602 GGAGCTCCAGCTTAGGG-----CCTGGCAGGAAGCTGGTTCATCAGGTGCTCCTCCTC 1549
Db 375 GlyAsnProGlyAlaAspGlyGlnProGlyAlaLys-GlyAlaThrGlyAlaPro----- 392
QY 1548 CTAGCACCCTCAGCTCCCTCGTATTTGGGAGGAGAACACCTGCTTCTCCGGTGGTAG 1489
Db 393 -----GlyIle---AlaGlyAlaProGlyPheProGlyAlaAr 404
QY 1488 AGGG-----AGGCCAGTGTGTAGGCGAGGATCTGCAGGCTGAGAGGTGAACCGGGT 1435
Db 404 gGlyProSerGlyProGlnGlyProSerGlyAlaProGlyProLys-GlyAsnSerGlyG 424
QY 1434 AGGGCGGTGAAGCTGTACCAGCGCCACACT----- 1403
Db 424 luProGlyAlaProGlyAsnLysGlyAspThrGlyAlaLysGlyGluProGlyProAlaG 444
QY 1402 -----GTGGACAGCATGTGGCAGCGCCAGCCCGCCAGCCAGGAAA 1366
Db 444 lyValGlnGlyProProGlyProAlaGlyGluGlyLysArgGlyAlaArgGlyGluP 464
QY 1365 GCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGA----- 1329
Db 464 roGlyProAlaGlyLeuProGlyProAlaGlyGluArgGlyAlaProGlySerArgGlyP 484
QY 1328 -----ATCGTGCACAGCCCGCTCCATGACCCAGAGAGA----- 1296
Db 484 heProGlyAlaAspGlyIleAlaGlyProLysGlyProProGlyGluArgGlySerProG 504
QY 1295 -----AGACCGGAGATGGCGACTG 1274
Db 504 lyAlaValGlyProLysGlySerProGlyGluAlaGlyArgProGlyGlyAlaGlyLeuP 524

QY 1273 CAGGAACA-----GCCCAAGCTGCCCATCCGAAAGCCCTTCATCATAGTG 1229
Db 524 roGlyAlaLysGlyLeuThrGlySerProGlySerProGlyPro-----AspGlyL 541
QY 1228 TCTCCGGGCTCGGTCCCGGCTCAGCTCTGGGCGAGCCCTGTACAGCCCTCGCCAC 1169
Db 541 ystThrGlyProProGlyProAlaGlyGlnAspGlyArgProGlyProAlaGlyProPro- 560
QY 1168 GAAATCCGTGTAAACAGCGTGAAGGTCTAGTGCCATCCAGCTGCACAGCTCAGCCAC 1109
Db 561 -----GlyA 562
QY 1108 GAAGACCCCGCAGGCTCGGGGATGCGGCGAGCACACCTGGTCGAGCCGGGAGAGCAG 1049
Db 562 laArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAlaGlyGluProG 582
QY 1048 GCGCCCGAGTTCC-----GGAAAGCCAGCGCGGCGCATGGACAGCAGTGGGCGACAA 992
Db 582 lyLysProGlyGluArgGlyAlaProGlyProProGly-----AlaValGlyAlaAlaG 600
QY 991 GGAGGGGCGCAGACGCCCTTCTGCTGGCTCGGTGGGCGCCAGCGCTGCTCCTCAGCCAC 932
Db 600 lyLysAspGlyGluAla-----GlyAlaGlnGlyProProGlyProT 614
QY 931 CAGCAGTGTGGCTGTACGCGAGTGAAGAGTGAAGGTGACGAGCGGCAAGAGAGCAGCT- 873
Db 614 hr-----GlyProAlaGlyGluArgGlyGluG 623
QY 872 -----CCTCTGGTGGTCCAGGT-----AGGGG-----CCAGGGCAGTGGTCCAGTCAAT 824
Db 623 lnglyProAlaGlyAlaProGlyPheGlnGlyLeuProGlyProAlaGlyProPro----- 641
QY 823 GCGAGCAGGAGGTAGCCCGCAGCAGCCCGCCAGACTGATCATGAAGCATAGACAGATA 764
Db 642 -----GlyGluAlaGlyLysProGlyGlu-----GlnGlyV 652
QY 763 GCGCTGGCGCAGTGGT-----CGGGTCCC-----GGAGAGGTGAGAGAGGCGGCTC 713
Db 652 alProGlyAsnAlaGlyAlaProGlyProAlaGlyAlaArgGlyGluArg-----GlyPheP 671
QY 712 CAGTGGAGTGAAGCAGCAGCTGGCCACAGAAAGTCCAGCAGCCCGCCAGCAGGATGAGCAG 653
Db 671 roGlyGluArgGlyValGlnGlyProProGlyProGlnGlyProArgGlyAlaAsnGlyA 691
QY 652 TGCCA-----GCTCCAGGGGCTGGATCCGGGCA 623
Db 691 laProGlyAsnAspGlyAlaLysGlyAspAlaGlyAlaProGlyAla-----ProGlyA 709
QY 622 CAGCAGCCCTGCTAGCCAGCGCGCTTGGATGAGAAAGAGGTGACAGCAGGATGCCCA- 564
Db 709 snGluGlyProProGlyLeuGlyMetProGlyGluArgGlyAlaAlaGlyLeuProG 729
QY 563 -----AGGACAGTGGCCAGATGAAGGGCGCGCGGCGC----- 531
Db 729 lyAlaLysGlyAspArgGlyAspProGlyProLysGlyAlaAspGlyAlaProGlyLysA 749
QY 530 -----CATAGCGTCCACCCAGTGGTC 509
Db 749 spGlyLeuArgGlyLeuThrGlyProGlyProGlyProGlyProAlaGlyAlaProGly- 768
QY 508 ACTGCTCAGCCTAGGAGCGGACACAGCAGCCAGCCCGCCAGCA-----CTGACCAANTGCC 455
Db 769 -----AspLysGlyGluAlaGlyProProGlyProAlaGlyProThrGlyAlaArgG 786
QY 454 CAGCACCATGTGTATGAATCTCTCTACCCCGCTTCCAGCAGCAGCAGCGCGCACATA 395
Db 786 lyAlaProGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGly----- 804
QY 394 GGTGATGCTCGGCGCAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAAGAGAG 335
Db 805 -----ProProGly-----AlaAspGlyGlnProGlyAlaL 815

Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
 [15]
 RP VARIANT OI-II ARG-569.
 RX MEDLINE-87222295; PubMed-3108247;
 RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
 RT "lethal perinatal osteogenesis imperfecta due to the substitution of
 RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I
 RT collagen.";
 RL J. Biol. Chem. 262:7021-7027(1987).
 RN [16]
 RP VARIANT OI-II CYS-926.
 RX MEDLINE-88033031; PubMed-3667599;
 RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
 RT "A point mutation in a type I procollagen gene converts glycine 748
 RT of the alpha 1 chain to cysteine and destabilizes the triple helix in
 RT a lethal variant of osteogenesis imperfecta.";
 RL J. Biol. Chem. 262:14737-14744(1987).
 RN [17]
 RP VARIANT OI-II ARG-842.
 RX MEDLINE-88298828; PubMed-3403550;
 RA Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;
 RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
 RT chain in lethal perinatal osteogenesis imperfecta. Demonstration of
 RT the peptide defect by in vitro expression of the mutant cDNA.";
 RL J. Biol. Chem. 263:11627-11630(1988).
 RN [18]
 RP VARIANT OI CYS-1195.
 RX MEDLINE-89218628; PubMed-3244312;
 RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
 RT "A cysteine for glycine substitution at position 1017 in an alpha
 RT 1(I) chain of type I collagen in a patient with mild dominantly
 RT inherited osteogenesis imperfecta.";
 RL Mol. Biol. Med. 5:197-207(1988).
 RN [19]
 RP VARIANT OI-II VAL-434.
 RX MEDLINE-89255493; PubMed-2470760;
 RA Patterson E., Smiley E., Bonadio J.;
 RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
 RT mutation.";
 RL J. Biol. Chem. 264:10083-10087(1989).
 RN [20]
 RP VARIANT OI-IV SER-1010.
 RX MEDLINE-89308591; PubMed-2745420;
 RA Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koepflin D.A.;
 RT "Osteogenesis imperfecta type IV. Detection of a point mutation in
 RT one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
 RL J. Biol. Chem. 264:11893-11900(1989).
 RN [21]
 RP VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
 RX MEDLINE-89380165; PubMed-2777764;
 RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
 RT "Characterization of point mutations in the collagen COL1A1 and
 RT COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
 RL J. Biol. Chem. 264:15809-15812(1989).
 RN [22]
 RP VARIANT OI SER-1022.
 RX MEDLINE-90062068; PubMed-2511192;
 RA Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
 RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe
 RT variant of osteogenesis imperfecta minimally destabilizes the triple
 RT helix of type I procollagen. The effects of glycine substitutions on
 RT thermal stability are either position of amino acid specific.";
 RL J. Biol. Chem. 264:19694-19699(1989).
 RN [23]
 RP VARIANT OI-II CYS-1082.
 RX MEDLINE-89109573; PubMed-2913053;
 RA Constantinou C.D., Nielsen K.B., Prockop D.J.;
 RT "A lethal variant of osteogenesis imperfecta has a single base
 RT mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
 RT chain of type I procollagen. The asymptomatic mother has an
 RT unidentified mutation producing an overmodified and unstable type I
 RT procollagen.";
 RL J. Clin. Invest. 83:574-584(1989).

[24]
 RN VARIANTS OI CYS-272; CYS-704 AND CYS-896.
 RP MEDLINE-90093113; PubMed-2794057;
 RA Starmen B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
 RA Weiss L., Graham J.M., Byers P.H.;
 RT "Osteogenesis imperfecta. The position of substitution for glycine by
 RT cysteine in the triple helical domain of the pro alpha 1(I) chains of
 RT type I collagen determines the clinical phenotype.";
 RL J. Clin. Invest. 84:1206-1214(1989).
 RN [25]
 RP VARIANT OI-II CYS-422.
 Alignment Scores:
 Pred No.: 9.22e-09 Length: 1464
 Score: 343.00 Matches: 266
 Percent Similarity: 32.32% Conservative: 42
 Best Local Similarity: 27.91% Mismatches: 331
 Query Match: 5.51% Indels: 315
 DB: 1 Gaps: 55
 US-09-759-143-110 (1-3410) x CALL_HUMAN (1-1464)
 Qy 2353 AGATGGGAAACAGGTGACTGAGTTTATTCAGCTCCCAAAACCCCT-----TCTCTA 2300
 Db 131 ArgAspGlyIleProGlyGlnProGlyLeuProGlyProGlyProGlyProGlyPro 150
 Qy 2299 GGTGTGTCCTCACTAGGAGGCTAGCTGTTA---ACCTGAGCCTGGGTAATCCACTGCA 2243
 Db 151 GlyProProGlyLeuGlyGlyAsnPheAlaProGlnLeuSerTyrGlyTyrAspGluLys 170
 Qy 2242 GAGTCCCGCATTCCTGAGGAGGCTTCTGGCCCTCCCTGTATTAAGTCCAGACTGAA 2183
 Db 171 SerThrGlyGlyIleSerValProGlyProMetGly----- 182
 Qy 2182 ACCCCCTTGAAGGCCTCCAGTCAGGAGCCCTAGAGACTGGGAGAGAGAGAGGAGCAGC 2123
 Db 183 -----ProSerGlyPro-----ArgGlyLeu 189
 Qy 2122 CCCAGAGCCCGAGCTGTGACAGTACGACCTCAGCAGCAGGAGGAGGAGGAGGAGGAGC 2063
 Db 190 ProGlyProProGlyAlaPro---GlyProGlnGlyPheGlnGlyProGlyGluPro 208
 Qy 2062 CATTACTTTCGAGCAACAAACTGGCGCCAGCCGCGGAGCCCATGGGCTAACAGG 2003
 Db 209 GlyCluProGlyAlaSerGlyProMetGlyProArgGlyProGlyProGlyProGlyLys 228
 Qy 2002 AGCGGGAGCTGGGA-----CCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1967
 Db 229 AsnGlyAspAspGlyGlyAlaGlyLysProGlyArgProGlyGlyArgGlyProGly 248
 Qy 1966 -----CCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1928
 Db 249 ProGlnGlyAlaArgGlyLeuProGlyThrAlaGly-----LeuProGly 263
 Qy 1927 CAAGTCGCTCTGTGCAAAATACCTACCTGTGTAGCAAAAGTAAATGGCAGAGAGGAGGAG 1868
 Db 264 MetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaGlyAlaGlyProAla 283
 Qy 1867 TCGGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1808
 Db 284 GlyProLysGlyGluProGlySerProGly-----GluAsnGlyAlaProGly 299
 Qy 1807 CAGGGATGGGGCCA-----CCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1757
 Db 300 Gln--MetGlyProArgGlyLeuProGlyGlyArg-----ProGlyAlaProG 317
 Qy 1756 GTCCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1718
 Db 317 lyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAlaGlyProGlyProGlyPro 337
 Qy 1717 GGTGGGCTACCCACCACACAC-----CGTAGGAGAGAGAGAGAGAGAGAGAGAGAG 1680
 Db 337 hrGlyProAlaGlyProProGlyPheProGlyAlaValGlyAlaGlyGlyGlyGly 357

QY	1802	ATGGGGCCACCTGGG----	ACAGCAGGAAGGCACACTAT	CCAGGATGGCGAGTCCAGCAGA	1746
DB					
QY	768	GluclyProGlyGluLysGlyGlyGlnGlyProGlyProGlyProGlnGlyProIleGly	787		
QY	1745	TGCCCCGGCCGGNA-----			1731
DB					
QY	788	TyrProGlyProArgGlyValLysGlyAlaAspGlyIleArgGlyLeuLysGlyThrLys	807		
QY	1731	-----			1731
DB					
QY	808	GlyGluLysGlyGluAspGlyPheProGlyPheLysGlyAspMetGlyIleLysGlyAsp	827		
QY	1730	-----	CCACCTCGCTCGGTGGGCTCACCCACACACACGTACGGAGA	1686	
DB					
QY	828	ArgGlyGluIleGlyProProGlyProArg--GlyGlu-----	839		
QY	1685	CATCACAGGCAGAGGCCCGCAGACAGCGGGTGGAGGTGGGAGCCCACTGCCCTCCAG	1626		
DB					
QY	840	--AspGlyProGluGlyProLysGlyArgGlyProAsnGlyAspProGlyProLeuG	859		
QY	1625	CACCCAGGTGCATTAGGGAAG--GGAGCTCCAGGCTTAGGG-----	CCTGGCAGGA	1575	
DB					
QY	859	lyProProGlyGluLysGlyLysLeuGlyValProGlyLeuProGlyTyrProGlyArgG	879		
QY	1574	AGCTGTGTCATCAGCGTCTCTACTCTAGCACCTCCAGTGTCCCTCGGTATT-----	1521		
DB					
QY	879	ln-----GlyProLysGlySer-IleGlyPheProGly	889		
QY	1520	TGGGCAGGAACACCTGCTTCTCCCGTGTAGAGGAGGCCAGTGTGTAGGCAGAGTCT	1461		
DB					
QY	890	PheProGlyAlaAsnGlyGluLysGlyArgGlyThrPro-----GlyLys	905		
QY	1460	GCAGGCTGAGAGGTGAACCCGGTGAGCGCGGTGAAGCTGTCAACGCGGCACACTGT	1401		
DB					
QY	906	ProGly-----			907
QY	1400	GGGACAGGCATGTGGCACCGGAGCCACAGGAAAGCTGCCACACTGGCCAAATAGACTG	1341		
DB					
QY	908	-----ProArgGlyGln-----	911		
QY	1340	CTCAGTGCCGAATCGCTGCACCCAGCGCGTCCATGACCAGAGA-----	1296		
DB					
QY	912	-----ArgGlyProThrGlyProArgGlyGluArgGlyProArgGlyIle	926		
QY	1295	-----AGACCAGG---AGATGGGCACTGCAGGAACAGCCCA---GGTGGCCATCC	1248		
DB					
QY	927	ThrGlyLysProGlyProLysGlyAsnSerGlyIleAspGlyProAlaGlyProProGly	946		
QY	1247	GAACGCTTCATAGTGTCTCCGGGCTCGGTGCCGGTCACTCTGGGCAGCGCT	1188		
DB					
QY	947	GluArg-----GlyProAsnGlyProGlnGlyProThrGlyPhePro	960		
QY	1187	GGTACAGCCCTCGCCCAAGAAATCCGTGTAAACAGCGTGAAGTCAATGAGTGCCA---	1131		
DB					
QY	961	GlyProLysGlyProProGlyProProGlyLysAspGly-----LeuProGly	976		
QY	1130	-----TCCAGCTGCACAGCTCAAGCCACGAGAGCC	1101		
DB					
QY	977	HisProGlyGlnArgGlyGluThrGlyPheGlnGlyLysThrGlyProProGlyProPro	996		
QY	1100	GGCGCAGGTGCGGGCATGCGGCAGCACA-----GCTGTGCAGCGCGGAGCAGGG	1047		
DB					
QY	997	GlyVal---ValGlyProGlnGlyProThrGlyGluThrGlyProMetGlyGluArgGly	1015		
QY	1046	CGCCCGAGTTCGGAAAGCCAAAGCGG-----CCCGCATGCAGACAGTGGG	999		
DB					
QY	1016	HisProGlyProProGlyProProGlyGlnGlnGlyLeuProGlyLeuAlaGlyLysGlu	1035		
QY	998	GGGACAAGGGGGCCACAGCCCTTCTGCTGGTGGTGGGCCCGCCAGCG---CTGCCT	942		
DB					
QY	1036	GlyThrLysGlyAspPro-----GlyProAlaGlyLeuPro	1047		
DB					

QY	941	CCTCAG-----	CCACACAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATCAGCGGTGAGCA	888
Db	1048	GlyLysAspGlyProPro-----		
QY	887	GGCCAAAGAGGCACCT-----		---GlyLeuArgGlyPhePro
Db	1060	GlyAspArgGlyLeuProGlyProValGlyAlaLeuGlyLeuLysGlyAsnGluGlyPro		---CCT
QY	869	CCTGGGTGCCAGGT-----	AGGGGCCAGGGCAGCTGCTGCCAGTCAATGGCAGCA	816
Db	1080	ProGlyProProGlyProAlaGlySerProGlyGluArgGlyProAla-----		---GlyAla
QY	815	GGAGGTAGCCAGCAGCCCCCAGACTGATCATGAAGCATAGACAGTAGTGGCTGGC		756
Db	1098	AlaGlyProIleGlyIlePro-----		---GlyArgProGly
QY	755	GACAGTGGTCCGGGTCCCGAAGAGGTCTCAGAGAGCGGCCCTCCAGTGGAGTGAAGCACA		696
Db	1109	ProGlnGlyPro-----		---ProGlyProAlaGlyGluLysGlyAla
QY	695	CCTGGCCACAGAAGTCCAGCAGGCCACCCAGGATGAGCAGTGCACAGTCCAGGGGCC		636
Db	1122	ProGlyGluLysGlyProGlnGlyProAlaGlyArgAsp-----		---GlyLeuGlnGlyPr
QY	635	TGGAGTCCGGGCACAGCAGCCCTGCTAGCAGCGCCCTTGGGATGAG-----		587
Db	1139	o-----ValGlyLeuProGlyProAlaGlyProValGlyProGlyGluAspGlyAsPly		1158
QY	586	-----AAAGAGCTCAGCAGGATCCCCAAGCAGCAGTGCACAGATGAAGGCC		540
Db	1158	sGlyGluIleGlyGluProGlyGlnLysGlySerLysGlyAspLysGlyGluGlnGlyPr		1178
QY	539	GGCGCGGCATAGTCGCACGCCAGTGGTCACTGGCTGAGCCTAGGA-----		CGC
Db	1178	oProGly-ProThrGlyPro-----		---GlnGlyProIleGlyGlnPro-GlyProSerGlyAla
QY	488	GGACACAGACAGGCCAGCAGCTGACCAAA-----		459
Db	1197	AspGlyGluProGlyProArgGlyGlnGlnGlyLeuPheGlyGlnLysGlyAspGluGly		1216
QY	458	-----TGCCAGCAGCATGGTCACTGAACCTCT-----		---CCTCTACCCCA
Db	1217	ProArgGlyPheProGlyProProGlyProValGlyLeuGlnGlyLeuProGlyPro		1236
QY	419	CTTCCAGCAGCAGCGGCACATAGGTGATGCTGCGGCCAAACACACCTCCAGGCCAA		360
Db	1237	GlyGluLysGlyGluThrGlyAspValGlyGlnMetGlyProProGlyProProGlyPro		1256
QY	359	AGG---TTAGCAGGTGACCAGCAGACTGGCTTTCCGGTGGCGCAGCAGGC-----		GGC
Db	1257	ArgGlyProSerGlyAlaProGlyAlaAspGlyProGlnGlyProProGlyGlyIleGly		1276
QY	305	TCACCCACAGCCTCTGGACCATAGTGGG-----		---CCAGCGGGTAGGCTCAGG
Db	1277	AsnProGlyAlaValGly-GluLysGlyGluProGlyGluAlaGlyGluProGlyProSe		1296
QY	257	GGCGCGTTCAGGCATCCA-----		---GAACCTGCTGCTCCGC-----TCTGCTCC
Db	1296	rGlyArgSerGlyProProGlyProLysGlyGluArgGlyGluLysGlyGluSerGlyPr		1316
QY	212	AGAAGCTCGCGCTCTCTCTCTGCTGTCGCG-----		---CACTGCCTAGGAA
Db	1316	oSerGlyAlaAlaGlyProProGlyProLysGlyProGlyAspAspGlyProLysGly		1336
QY	167	TCAGCCAGCGGCCCATTTCTGCCAGCCCTTTGGT-----		---GCCGGTCCAGCTTC
Db	1336	ySerProGly-----		---ProValGlyPheProGlyAspProGlyPro
QY	119	TCAGCCCATGCTCAACACCTGCTGCTGGGGCACCTCAGTGGGGACACGCTCATCACT		60
Db	1351	yGluPro-----		---Gly-ProAlaGlyGlnAspGlyProProGlyAspLysGlyAspG
QY	59	CAGATCTGGCGAGGCGCGCGCTGTCAACCCGGA		25

CC	DR	EMBL; M92642; AAA58427.1; -			
DR	EMBL; S57132; AAB25797.1; -				
DR	PIR; S23810; S23810.				
DR	Genew; HGNC:2193; COL16A1.				
DR	MIW; 120326; -				
DR	InterPro; IPR000087; Collagen.				
DR	InterPro; IPR003129; TSPN.				
DR	Pfam; PF01391; Collagen; 18.				
DR	Pfam; PF02210; TSPN; 1.				
DR	SMART; SM00210; TSPN; 1.				
DR	Extracellular matrix; Connective tissue; Collagen; Hydroxylation;				
KW	Repeat; Signal.				
KW	SIGNAL	1	21		POTENTIAL.
FT	CHAIN	22	1603		COLLAGEN ALPHA 1(XVI) CHAIN.
ET	DOMAIN	22	374		NONHELICAL REGION 10 (NC10).

Db 476 lyAsp-----LysG 479
QY 2646 CTAGCAATGGATTCCCTTCTACTTTGTTAAATAAATAGTTAAATATTTAAATGCCCTGTG 2587
Db 479 lySerSerGlyIlePro-----484
QY 2586 TCTCTGTGATGCCAACAGAGACCAAGACAGGCCACATCCTGATAAAAGSTAAAGAGGGGG 2527
Db 485 -----GlyLysGluGlyProGlyGlyLys-----ProGlyLysProGlyV 498
QY 2526 TGGATCAGCAAAAAGACAGTCTGTGGCTGAGGGGACCTGGTTCCTGTGTGCCCT 2467
Db 498 alLysGlyLys-----GlyAsp-----504
QY 2466 CAGGACTCTCCCTACAAATAAGTCATATGTTCAAAATCCCATGGAGCTGTTTCATCC 2407
Db 505 -----ProCysGluValCysProThrL 512
QY 2406 TAGAACTCCCATGCAAGAGCTACATTAAACGAAGCTGAGGCTTAAGGGCTTAGAGATG 2347
Db 512 euProGluGlyPheGlnAsnPhe-----ValGlyLeuProGlyLysProGlyProLysG 530
QY 2346 GGAACACAGGTGACTGATTTATCA-----G 2320
Db 530 lyGluProGlyAspProValArgAlaArgGlyAspProGlyIleGlnGlyLysGlyG 550
QY 2319 CTCCTCAAAAACCTTCTCTAGGTGTCTCACTAGGAGCTAGCTGTTAACTTACCTGAGC- 2261
Db 550 luLysGlyGluProCysLeuSerCysSerValValGlyAlaGlnHisLeuValSerS 570
QY 2260 -----CTGGGTAAATCCA-----2249
Db 570 erThrGlyAlaSerGlyAspValGlySerProGlyPheGlyLeuProGlyLeuProGlyA 590
QY 2248 -----CCTGCAG 2242
Db 590 rgAlaGlyValProGlyLeuLysGlyGluLysGlyAsnPheGlyGluAlaGlyProAlaG 610
QY 2241 AGTCCCG-----CATTCAGTGCATGGAGCCCTCTTGGCCCTC-----2204
Db 610 lySerProGlyProGlyProValGlyProAlaGlyIleLysGlyAlaLysGlyGluP 630
QY 2203 -----CCTGTATAAGTCCAGACTGAACCCCTTGGAGGCTCCACTCAG 2158
Db 630 roCysGluProCysProAla-----LeuSerAsnLeuGlnAspG 643
QY 2157 GCAGC-----CCTAGACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2116
Db 643 lyAspValArgValAlaLeuProGlyProSerGlyGlu-----LysGlyGluProGlyP 662
QY 2115 CCCAGGTGTGACAGTACGCACCTCAGCAGCAGAGGTGGCAGCAGAGAGAGAGAGAGAGAG 2056
Db 662 roProGlyPheGlyLeu-----ProGlyLysGlnGlyLysAlaGlyGluArgGlyLeuL 680
QY 2055 TTGSCAGACACAGAACTGGCGCCAGCCCGGAGCC-----ATGGGGCTAACAGGA 2002
Db 680 ysGlyGlnLysGlyAsp-AlaGlyAsnProGlyAspProGlyThrProGlyThrThrGly 699
QY 2001 GCGGG-----AGCTGGGACCA-----GTGAGCAGGCGCTCCACCCCAATGTGTCGAA 1951
Db 700 ArgProGlyLeuSerGlyGluProGlyValGln-GlyProAlaGlyProLysGlyGluL 719
QY 1950 GTTTTCTACCTGATTTGGCCAAAGTCTCTTCTCAANTACTACCTGTGTAGCAAG 1891
Db 719 sGlyAspGlyCysThrAlaCysProSerLeuGlnGlyThrValThrAspMetAlaGlyAr 739
QY 1890 TAAATGGGACACAGCCAGGCTCGGGCAGACACCATATAGGACGTGACAGTGGGCG 1831
Db 739 gProGlyGlnProGlyProLysGlyGluGlnGlyProGlyValGlyArgProGlyL 759
QY 1830 AGCTGGACAATGAGGCCATTAACAGAGGATGGGGCCACCTGGGACAGCAGGA---AGGCA 1774
Db 759 sProGlyGlnProGlyLeuProGlyValGlnGlyProGlyLeuLysGlyValGlnG 779

QY 1773 CTATCCAGGATGCCAGGTCCAGGCAGA-----1746
Db 779 yGluProGly-----ProProGlyArgGlyValGlnGlyProGlnGlyGluProGlyAl 797
QY 1745 -----TGCCCGGCCCGGAA-----CCACCTTGGCCT-----1719
Db 797 aProGlyLeuProGlyIleGlnGlyLeuProGlyProArgGlyProGlyProThrG 817
QY 1718 -----CGTGGGTCAACCCACACAGCTAGGAGACATCAGCAGC-----1676
Db 817 yGluLysGlyAlaGlnGlySerProGlyValLys-GlyAlaThrGlyProValGlyProP 837
QY 1675 -----A 1675
Db 837 rGlyAlaSerValSerGlyProGlyProGlyArgAspGlyGlnGlnGlyGlnThrGlyLeuA 857
QY 1674 GAGGC-----CCCGCAGAGCGGTGGAGTGGGAGCAGCCACTGCCTCCAGCACCACG 1618
Db 857 rgGlyThrProGlyGluLysGlyProArg-GlyGluLysGlyGluProGlyGluCysSer 876
QY 1617 TGTCCATTAGG-----AAGGAGCTCCAGCTTAGGCGCT 1582
Db 877 CysProSerGlnGlyAspLeuIlePheSerGlyMetProGlyAlaProGlyLeuTrpMet 896
QY 1581 GGCAGGAAGCTGTCTATCAGGCTGTCTCTAGCTAGCACCTCCAGTGTCCCTCGGTAT 1522
Db 897 Gly-SerSerTrpGlnProGly-----ProGlnGlyProGlyI 910
QY 1521 TTGGGAGGAAACCTCTCTCCCGGTGTAGAGGAGCGGCTGTGTAGGCGAGGATC 1462
Db 910 e-----ProGlyProGlyProGlyValProGlyLeuGlnGlyVal-----925
QY 1461 TCAGGGCTGAGAGGTGAACCCGGTGGGCGGTGAAGCTGTCCACACGCG-----1409
Db 926 --ProGlyAsnAsnGlyLeuProGlyGlnProGlyLeuThrAlaGluLeuGlySerLeuP 945
QY 1408 -----CACACTGTGGGACAGGCTGTGGCAGCCGCGCAGCAGCAGGAGAACTGCC 1360
Db 945 rolleGluGlnHisLeuLeuLysSerIleCysGly-----956
QY 1359 ACATGGGCAATAGACTCTCGAGTGCCTGCTGCACAGCGCGGTCCATGACCAGA 1300
Db 957 -----AspCysValGlnGlyGlnArgAlaHisProGlyTyrLeuValGluL 972
QY 1299 GAGAGACAGGAGATGCGGCACATGCAGGAACACCCCGAGCTGCCATCCGACGCGCT 1240
Db 972 ysGlyGluLysGlyAspGlnGlyIle-----ProGlyVal- 983
QY 1239 TCATCATAGTGTCTCCGGCCCTCGGTGCGGCTCAGCTCTGGGCGACCGCTGTGTACAGC 1180
Db 984 -----ProGlyLeuAspAsnCysAlaGlnCysPheLeuSerLeuGluArg 999
QY 1179 CCTCGCCCAACAAATCCGTGTAAACAGCGGTGAAGTGCATGAGTGCATCCAGC-TGCA 1121
Db 999 roArgAlaGluGluAlaArgGlyAspAsnSerGluGly-----AspProGlyCysV 1016
QY 1120 CAGCTCAGCA-----CGAAGACCCGCGCAGGCT 1091
Db 1016 alGlySerProGlyLeuProGlyProGlyLeuProGlyGlnArgGlyGluGluGlyP 1036
QY 1090 GCGGGCATGCGGACAGCAGCTGTGTGCGGCGGGAAGCAGCGCGCCAGCT-----1038
Db 1036 roProGlyMetArgLysSerProGlyProGlyProGlyProLysGlyProGlyPhePro 1056
QY 1037 -----TCCGGAAGCAAGCGGCGCGCATGGACAGCAGTGGGCGCAGAGGAGGGG 983
Db 1056 lyAlaValGlySerProGlyLeuProGlyLeuGlnGly-----GluArgGlyL 1072
QY 982 CGACAGCCCTTCTGCTGCTGGTGGGCGCCAGCGCTCCCTCCCTCAGCA-----933
Db 1072 euThrGlyLeuThrGlyAspLysGlyGluProGlyProGlyGlnProGlyTyrProG 1092

Qy 932 -----CCAGCAGTGTGGCTGCTACGAGGTGAGGAAGATGAG 896
 Db 1092 lyAlaThrGlyProProGlyLeuProGlyLeuGlyGluArg-----GlyTyrThrG 1110
 Qy 895 GGTGAGCAGGCCAAAGAGGACTCTCT-----CTGGGTGCG 860
 Db 1110 lySerAlaGlyGluGlyGluProGlyProGlySerGlyGluLeuProGlyProP 1130
 Qy 859 CAGGT-----AGGGGGCCAGGCGACTGGTCTCCAGTCAATGCC 821
 Db 1130 roGlyProAlaGlyProArgGlyGluArgGlyPro----- 1141
 Qy 820 AGGCAGGAGGTAGCCAGGAGCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGC 761
 Db 1142 -----GlnGlyAsnSerGlyGluGlyGlnP-----GlnGlyPheGlnGlyGlnP 1157
 Qy 760 CTGGCGACAGTGTGGGT-----CCCGGAAGAGGTGACAGAGCAGGCG 716
 Db 1157 roGlyPheThrGlyProThrGlySerProGlyPheProGlyLysValGlySerProGlyP 1177
 Qy 715 CTCCA-----GTGAGCTGAAGACACCTGCGCCAGCAAGTCCAGAGCCCGCCAGCGC 665
 Db 1177 roProGlyProGlnAlaGlyLysGlySerGluGlyLeuArgGlyProSerGly---LeuP 1196
 Qy 664 CAGGATCAGCAGTCCAGCTCCAGGGGCTGGGATCGGGGCACAGCAGCCCTCTAGCCA 605
 Db 1196 roGlySerProGlyProProGlyProGlyGlnGly-----Pro 1210
 Qy 604 GCCGCG-----CCTTGGGATGAAGAGGCTCAGCAGGATGCCAAGAGCAGTCCCA 551
 Db 1211 AlaGlyLeuAspGlyLeuAspGlyLysProlGlyLeuArgGly----- 1227
 Qy 550 GATGAAGGGCGGGCGG-----GCCATAGCTGCCAGCGCTGCTGCTGCTGAGCC 497
 Db 1228 AspProGlyProAlaGlyProProGlyLeuMetGlyProGlyPheLysGly---LysT 1247
 Qy 496 TAGAGCGGGACACAGCAGCGCCAGCTCCAGGCTGAGCACTGCGGCACAGCAGCCATGCTGTA 437
 Db 1247 hrGlyHisProGlyLeuProGlyProLysGlyAsp---CysGlyLysProGly----- 1263
 Qy 436 CTTCTCTCTACCCCA-----CTTCAGCAGCAGCAGCGCG----- 402
 Db 1264 -----ProProGlySerThrGlyArgProGlyAlaGluGlyGluProGlyAlaM 1280
 Qy 401 -----GCACATAGGTGATGCTGCGCGCCCAACACAC 371
 Db 1280 etGlyProGlnArgProGlyProProGlyHis-----ValGlyProProGlyP 1297
 Qy 370 CTCAGCGCCAAAGTTAGCAGGTGACACAGCAGCTGGCTTCCGGT----- 321
 Db 1297 roProGlyGln-----ProGlyProAlaGlyLysSerAlaValGlyLeuL 1312
 Qy 320 -----GCCGAGCAGGCGGCTCACCCACAG 296
 Db 1312 ysGlyAspArgGlyAlaThrGlyGluArgGlyLeuAlaGlyLeuProGlyGlnProGlyP 1332
 Qy 295 CCTGTGGACCATAGTGG-----CCAGCGGGTAGGCTCAGGGGCGGTTC 248
 Db 1332 roProGly-HisProGlyProProGlyGluProGlyThrAspGlyAlaAlaGlyLysGlu 1351
 Qy 247 GGCACCTCAGAACTGCTCGCTCGCTCTCCAGAGCTGGCGCTCTCTCTCTCTG 188
 Db 1352 GlyProProGlyLysGlnGlyPheTyrGlyProProGlyProLysGlyAspProGlyAla 1371
 Qy 187 TCCGCGCCAACTGCTAGGAATCAGCAGGCGCCATTTCTGCCAGCCCTTGTGGCGGT 128
 Db 1372 AlaGlyGln-----LysGlyGlnAlaGlyLysGlyArgAlaGlyMetProGlyGly 1389
 Qy 127 CCAGCTTCTCAGCCCATCAACACCTGCTGCTGCGGCGACCTCAGTGGGACAGCTC 68
 Db 1390 ProGlyLysSerGlySerMetGly-ProValGlyProGlyProAlaGlyGluArgL 1409
 Qy 67 TCATCACTCAGATCCTGGCGGAGGCGGCGCTCTCACCCGGA 25

Db 1409 yHisProGlyAlaProGlyProSerGly-----SerProGly 1421
 RESULT 15
 STP_SPIO
 ID STP_SPIO STANDARD; PRT; 525 AA.
 AC Q03411;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sucrose transport protein (Sucrose permease) (Sucrose-proton symporter).
 OS Spinacia oleracea (Spinach).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.
 OC NCBI_TaxID=3562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RX MEDLINE=93099843; PubMed=1464305;
 RA Riesmeier J.W., Willmitzer L., Frommer W.B.;
 RT "Isolation and characterization of a sucrose carrier cDNA from spinach by functional expression in yeast";
 RL EMBO J. 11:4705-4713(1992).
 CC -!- FUNCTION: RESPONSIBLE FOR THE TRANSPORT OF SUCROSE INTO THE CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM). CAN ALSO TRANSPORT MALTOSE AT A LESSER RATE.
 CC -!- PATHWAY: Sucrose metabolism.
 CC -!- SUBCELLULAR LOCATION: Inner membrane.
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC EMBL; X67125; CAA47604.1; -
 DR PIR; S28052; S28052.
 DR InterPro; IPR003662; sub_transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 DR TIGRFAMS; TIGR01301; GPH_sucrose; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
 KW Transmembrane; Transport; Sugar transport; Symport.
 FT DOMAIN 1 37 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 38 58 1 (POTENTIAL).
 FT TRANSMEM 72 92 2 (POTENTIAL).
 FT TRANSMEM 107 127 3 (POTENTIAL).
 FT TRANSMEM 145 165 4 (POTENTIAL).
 FT TRANSMEM 184 204 5 (POTENTIAL).
 FT TRANSMEM 230 250 6 (POTENTIAL).
 FT TRANSMEM 295 315 7 (POTENTIAL).
 FT TRANSMEM 338 358 8 (POTENTIAL).
 FT TRANSMEM 373 393 9 (POTENTIAL).
 FT TRANSMEM 422 442 10 (POTENTIAL).
 FT TRANSMEM 455 475 11 (POTENTIAL).
 FT TRANSMEM 488 508 12 (POTENTIAL).
 FT DOMAIN 509 525 CYTOPLASMIC (POTENTIAL).
 SQ SEQUENCE 525 AA; 54992 MW; 018347MD2CC1CC6 CRC64;
 Alignment Scores:
 Pred. No.: 5,25e-08 Length: 525
 Score: 327.00 Matches: 134
 Percent Similarity: 39.58% Conservative: 92
 Best Local Similarity: 23.47% Mismatches: 200
 Query Match: 5.10% Indels: 145
 DB: 1 Gaps: 21
 US-09-759-143-110 (1-3410) x STP_SPIO (1-525)

QY 310 CCTGCTGGCGCACCGGAAGCCACGACTCTTGCTGGTCAACCTGTCAACCTTTGGCCTCGA 369
||| ||||| ||||| ||| ||| ||| |||||
Db 25 ProThrThrProProGlu-AlaGluAlaThrLeuLysLysLeu-----GlyLeuVa 41

QY 370 GGTGTGTTGGCCGCAGCATCACTAT-----GTCCGCC 405
:::|||||||::: :::
Db 41 lAlaSerValAlaAlaGlyValGlnPheGlyTrpAlaLeuGlnLeuSerLeuLeuThrPr 61

QY 406 TCTGCTGCTGGAAGTGGGGTAGAGAGAAGTTTCATGACCATTGGTCTGGCATTTGGTCC 465
:::|||||||::: :::
Db 61 oTyrvValGlnLeuLeuGlyIleProHisThrTrpAlaAlaTyrlleTrpLeuCysGlyPr 81

QY 466 AGTGTGGCGCTGTCTGTGTCGCCGCTCCTAGGCTCAGCCAGTACCACCTGGCTGGACG 525
l::: |||||::: ||||| ||||| |||||
Db 81 oileSerGlyMetIleValGlnProLeuValGlyTyTrSerAspArgCysThrSerAr 101

QY 526 CTATGGCCGCGCGGCCCTTCANCTGGGCACTGCTTGGCATCTGGCATCTGTCAGCCTCT 585
g::::||| ||||| ||||| ||||| |||||
Db 101 gPheGlyArgArgPropHeIleAlaAlaGlyAlaAlaLeuValAlaValAlaValGl 121

QY 586 TCTCATCCCCAAGGCGCGCTGGCTAGCA-----GGGCTGTGTGCCGGATCCC----- 634
||| ||||| ||||| |||||
Db 121 yLeuIle-----GlyPheAlaAlaAspIleGlyAlaAlaSerGlyAspProThrGl 138

QY 635 -----AGGCCCTGGAGCTGGCACTGCTCATCTGGCGCTGGCGCTGGCGTCTGGACT 684
::: |||||::: ||||| ||||| |||||
Db 138 yAsnValAlaLysProArgAlaIleAlaValPheValValGlyPheTrpIleLeuAspVa 158

QY 685 CTGTGGCCAGGTGTCTTCACCTCCACCTGGAGGCCCTGTCTCTCACCTCTTCCGG---GA 741
||| ||||| ||||| ||||| |||||
Db 158 lAlaAsnAsnThrLeuGlnGlyProCysArgAlaLeuLeuAlaAspMetAlaAlaGlySe 178

QY 742 CCCGGACCACTGCGCCAGGCTTACTGTCTATGCTTCATCATCTGAGTATCTGGGGCTG 801
||| ||| :::::||| :::::|||
Db 178 rGlnThrLysThrArgTyrrAlaAsnAlaPhePheSerPheMetAlaLeuGlyAsnIl 198

QY 802 CTTGGGTACCTCTCTGCTGCTGCATTGAC-----TGGGACAC 837
||||| |||||
Db 198 eGlyGlyTyrrAlaAlaGlySertyrSerArgLeutyrrThrValPheProPhetrLysTh 218

QY 838 CAGTGCCTTGGCCCCCTACCTGGGCACCCAGGAGAGTGCTCTTTGGCTGCTCAC-- 895
l::: ||||| ||||| :::::|||
Db 218 rAlaAlcYaspValTyrrCysAlaAsnLeuLysSerCysPhePheIleSerIleThrLe 238

QY 896 -CTCATCTTCTCACCCTGCTAGCAGCACACTGCTGGTGGCTGAGGAGCAGCGCTGGG 954
||| ||||| ||||| ||||| |||||
Db 238 uLeulleValThrIleLeuAlaLeuSerValValLysGluArgGlnIleThrIleAs 258

QY 955 CCCACCGACCCACAAGGCTGTGGCCCCCTCTTGTGCGCCCCACTGCTGTCATG 1014
::: ||| ||| :::::|||
Db 258 pGluIleGlnGluGluAspLeuLysAsnArgAsnSerSerGlyCys----- 275

QY 1015 CCGGGCCCGCTTGGCTTTCGGAAACCTGGCGCCCTGCTTCCCGGCTGCACCACTGTG 1074
||||| ||||| ||||| ||||| |||||
Db 276 ---AlaArgLeuProPhe-----PheGlyGlnLeuilleGlyAlaLeuLysAspLeu-- 291

QY 1075 CTCGGCGATGCCCGGCACCTGCGCCCGCTTCTTGGTGGCTGAGCTGTGCAGCTGGATGC 1134
|||||::: |||||::: |||||:::
Db 292 -----ProLysProMetLeuilleLeuLeuLeuValThrAlaLeuAsnTrpIleAl 308

QY 1135 ACTCATGACCTTCACGCTGTTTTACAGGATTTCTGGGGAGGGGCTGTACCAGGCGT 1194
||| ||||| ||||| ||||| ||||| |||||
Db 308 atRpPheProPheLeuPheAspThrAspTrpMetGlyLysGluValTyrrGlyGlyTh 328

QY 1195 GCCCAGAGCTGACCGGGGCACCGAGGCCCGGAGACACTATGATGAAGCGCTTCGGATGGG 1254
::: |||||::: |||||
Db 328 r-----ValGlyGluGlyLysLeuTyrrAspGlnGlyValHisAlaGl 342

QY 1255 CAGCCTGGGCTGTCTCTGCAGTCCGCCCATCTCCCTGGCTCTCTCTGTGCTATGGACCG 1314
::: ||||| ||||| ||||| |||||
Db 342 yAlaLeuGlyLeuMetIleAsnSerValValLeuGlyValImeSerLeuSerIleGl 362

QY	1315	GCTGTGTCAGCGATTTCGC-----ACTCGAGCAGTCTATTTTGCCAGTGTGGCAGCAGTTT	1368
Db	362	yLeuAlaArgMetValGlyGlyAlaLysArgLeuTyrGlyValValAsnIleLeu--	381
QY	1369	CCCTGTGGCTGCCGGTGCACATGCCTGTCCACAGTGTGCCCTGGTGTGACAGCTTCAGC	1428
Db	382	-----AlaValCysLeuAlaMetThrVal---LeuValThrLysSerAl	395
QY	1429	CGCCCTCACCGGGTTTCACCTTCTCAGCCCTCGACATCCTGCCCTACACACTGGCGCTCCCT	1488
Db	395	aglu-----	396
QY	1489	CTACCACCGGAGAGCAGGTGTTCTTCGCCCAATACCGAGGGGACACTGGAGGTGCTAG	1548
Db	397	-HisPheArgAspSerHisHisIleMet-	405
QY	1549	CAGTGAGGACACGCTGTGATGACCGCTTCTGCCAGGCCCTTAAGCCTGGAGCTCCCTTCCC	1608
Db	405	-----	405
QY	1609	TAATGGACAGTGGGTGCTGAGGACAGTGGCCTGCTCCACCTCCACCCGCGCTCTGCGG	1668
Db	406	-----GlySerAlaValProProProProAla-----GI	416
QY	1669	GGCCTCTGCCTGTGATGTCCTCGTACCTGGTGGTGAGCGAGCCACCCAGGCCAGGCT	1728
Db	416	yValIysGlyAlaLeuAlaIlePheAlaValLeuGlyIleProLeuAlaIleThrPh	436
QY	1729	G-----GTTCCG-----GGCGGGGCATCTGCCT	1752
Db	436	eSerIleProPheAlaLeuAlaSerIlePheSerAlaSerSerGlySerGlyGlnGlyLe	456
QY	1753	GGACCTCGCCATCTCGATAGTGCCTTCCTGCTGCCAGTG-----	1795
Db	456	uSerLeuGlyValLeuAsnLeuAlaIleValProGlnMetPheValSerValThrse	476
QY	1796	-----GCCCATCTCCCTGTTTATGGCTGCCAT	1821
Db	476	rglyProTyrAspAlaMetPheGlyGlyGlyAsnLeuProAlaPheValGlyAlaVala	496
QY	1822	TGTCAGCTCAGCCAGTCTGCTCAGTGCTAT	1852
Db	496	IalaAlaThrAlaSerAlaValLeuSerPhe	506
RESULT 16			
CA12_MOUSE			
ID	CA12_MOUSE	STANDARD;	PRT; 1459 AA.
AC	P28481;		
DT	01-DEC-1992 (Rel. 24, Created)		
DT	01-DEC-1992 (Rel. 24, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Collagen alpha 1(I) chain precursor [Contains: Chondrocalcin].		
OS	COL2A1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=91358489; PubMed=1885613;		
RA	Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;		
RT	"Mouse type II collagen gene. Complete nucleotide sequence, exon		
RT	structure, and alternative splicing."		
RL	J. Biol. Chem. 266:16862-16869(1991).		
RN	[2]		
RP	SEQUENCE OF 1455-1459 FROM N.A.		
RX	MEDLINE=91274355; PubMed=2054384;		
RA	Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;		
RT	"Specific hybridization probes for mouse type I, II, III and IX		
RT	collagen mRNAs."		
RL	Biochim. Biophys. Acta 1089:241-243(1991).		
CC	-!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.		
CC	-!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.		

-|- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 -|- SIMILARITY: CONTAINS 1 VWFC DOMAIN.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M65161; AAA68100.1; -;
 EMBL; X57982; CAA41047.1; -;
 MGD; MGI:88452; Col2a1.
 InterPro; IPR000087; Collagen.
 InterPro; IPR000885; Fib_collagen_C.
 InterPro; IPR001007; VWFC_C.
 Pfam; PF00093; vwc; 1.
 Pfam; PF01391; Collagen; 18.
 Pfam; PF01410; COLFI; 1.
 ProDom; PD000007; Collagen; 3.
 ProDom; PD002078; Fib_collagen_C; 1.
 SMART; SM00038; COLFI; 1.
 SMART; SM00214; VWC; 1.
 PROSITE; PS01208; VWFC; 1.
 Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.
 SIGNAL 1 25
 PROPEP 26 153
 AMINO-TERMINAL PROPEPTIDE
 (BY SIMILARITY).
 CHAIN 154 1213
 PROPEP 1214 1459
 COLLAGEN ALPHA 1(II) CHAIN.
 CARBOXYL-TERMINAL PROPEPTIDE
 (CHONDROCALCIN).
 VWFC.
 TRIPLE-HELICAL REGION.
 NON-HELICAL REGION (C-TERMINAL).
 Q -> R (IN SHORT ISOFORM).
 MISSING (IN SHORT ISOFORM).
 SEQUENCE 1459 AA; 139154 MW; F6C84FA7C532E7F2 CRC64;

Alignment Scores:

Pred. No.: 5,23e-08 Length: 1459
 Score: 326.00 Matches: 293
 Percent Similarity: 31.97% Conservative: 59
 Best Local Similarity: 26.61% Mismatches: 407
 Query Match: 5.24% Indels: 344
 DB: 1 Gaps: 63

US-09-759-143-110 (1-3410) x CAL2_MOUSE (1-1459)

QY	2557	GGCCACATCTGATAAAAGGTAAAGGGGTGGATCAGCAAAAGACAGTGTGTGGC	2498
DB	32	GlySerCysLeuGlnAsnGlyClnArg	47
QY	2497	TGAGGGGACCTGGTCTTCTGTGTGT	2441
DB	48	LysProSerSerCysArgGileCysValCysAspThrGlyAsnValLeuCysAspAsp	67
QY	2440	ATATGTCAATCCCATGGAGGAGTGTTCATCTCTAGAACTCCCATGCAAGAG	2387
DB	68	IleCysGluAspPro	85
QY	2386	-----CTACATTAAACGAAGCTGCAGGTTTAAGGGGCTT	2354
DB	86	ProIleCysProAlaAspLeuAlaThrAlaSerGlyArgLysLeuGlyProLysGly	104
QY	2353	AGAGATGGAAACCAAGGTGACTGTTTATTCAGCTCCCAAAACCTCTCTAGGTGTG	2294
DB	105	GlnLysGlyGluProGlyAspIleArgAspGlyAlaPro	123
QY	2293	TCTCAACTAGGAGGCTAGTGTGTTAACTCCCTGGGTGTAATCCA	2237

DB	124	GlyGluProGly	137
QY	2236	CGCATTCAGTGCATGGAGCCCTTCTGGCCCTCCCTGATAAGTCCAGACTGAAACCC	2177
DB	138	Pro	144
QY	2176	TTGGAAGGCTCCAGTCAGGAGC	2147
DB	145	-----GlyLeuSerAlaGlyAsnPheAlaGlnMetAlaGlyGlyTyrAspGlu	162
QY	2146	GACTGGGAGAGAGAGGAGC	2123
DB	163	AlaGlyGlyAlaGlnMetGlyValMetGlnGlyProMetGlyProMetGlyProArgGly	182
QY	2122	CCCCAGCCCGAGCTGTGCAGCTACCTCAGCAGCAGCGGTGGCAGCAGAGCA	2063
DB	183	ProProGlyProAlaGlyAlaProGlyProGlnGlyPheGlnGlyAsnProGlyGluPro	202
QY	2062	CATTACTTTGGCAGCAACAGAACTGGCGGCCAGCCGCCATGGGGTCAACAGG	2003
DB	203	GlyGluProGlyValSerGlyProMetGlyProArgGlyProGlyProAlaGlyLys	222
QY	2002	AGCGGGAGCTGGGA	1967
DB	223	ProGlyAspAspGlyGluAlaGlyLysProGlyLysSerGlyGlyGluArgGlyLeuProGly	242
QY	1966	CCCCAATGTCTGGAAGTTTCTACGCTGAGTATTGGCCAGTGCCTCTTCAATAC	1907
DB	243	ProMetGlyAlaArgGlyPhePro	255
QY	1906	TACCTGTGTAGCAAACTAATGCGCAGCAGAC	1866
DB	256	-----ProGlyValLysGlyHisArgGlyTyrProGlyLeuAspGlyAlaLysGly	272
QY	1865	---CGCAGACACCATATAGGCTGACAGCTGCTGAGCTGGCAATGGAGCCCATAA	1809
DB	273	GluAlaGlyAlaProGlyValLysGlyGlySerGlySerProGlyGlyLeuAsnGlySerPro	292
QY	1808	ACAGGATGGGGCA	1767
DB	293	GlyProMetGlyProArgGlyLeuProGlyGlyArgGlyThrGlyProAlaGlyAla	312
QY	1766	---GGATGGGAGGTCCAGCAGATGCCCGGGCCCGAA	1717
DB	313	AlaGlyAlaArgGlyAsnAspGlyGlnProGlyProAlaGlyProGlyProValGly	332
QY	1716	---GTGGGTCAACCCACACCATCAGGACATCACAGCAGAGAGCC	1669
DB	333	ProAlaGlyGlyProGlyPheProGlyAlaProGlyAlaLysGlyGlyAlaGlyProThr	352
QY	1668	-----CCGAGAGCGGGTGGAGTGGGAGCAGCCACTGCTCCAGCACCC	1621
DB	353	GlyAlaArgGlyProGlyGlyAlaGlnGlySerArgGlyGlyProGlyAsnProGlySer	372
QY	1620	ACGTGTCCATTAGG	1570
DB	373	ProGlyProAlaGlyAlaSerGlyAsnProGlyThrAspGlyThrProGlyAlaLys-Gl	392
QY	1569	GTCAATCAGGCTGCTCAGTGTAGCACCTCCAGTGTCCCTCGGTATTGGCGCAGAAC	1510
DB	392	ySerAlaGlyAlaPro	402
QY	1509	ACCTGCTTCTCCCGTGGTAGAGGAGCCAGTGTGTAGGCA	1465
DB	402	aproglyPheProGlyProArgGlyProGlyProGlnGlyAlaThrGlyProLeuGly	422
QY	1464	ATCTGAGGCTGAGAGGTGAACCCGGTGGCGGCTGAAGCTGTCAACAGGC	1409
DB	422	yPro-LysGlyGlnAlaGlyGluProGlyAlaGlyPheLysGlyAspGlnGlyProL	442
QY	1408	---CACACTGTGGACAGGATGTGGCACCGGACAGGAGGAGCTCCACACTGGCC	1351
DB	442	ysGlyGluThrGlyProAlaGlyProGlnGly-AlaProGlyProAlaGlyGluGly	461

QY 1350 AAATAGACTGCTCGA-----GTCCGAATCGCTGCACACCGCGTCCATCACCAGAG 1297
 DB 462 LysArgGlyAlaAArgGlyGluProGlyGly-AlaGlyProGlyGlyProGlyGly 481
 QY 1296 AAGA-----CCAGGAGATGCGCGACTGCAGAACACGCCCCAGG 1258
 DB 481 gGlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaGly---ProLysG 500
 QY 1257 CTGCCCCATCGAAGCCCTTCATCATAGTGTCTCCGGGCTCCG----- 1215
 DB 500 yAlaProGlyGluArgGlyProSerGlyLeuAlaGlyProLysGlyAlaAsnGlyAspPr 520
 QY 1214 ----TGCCGGCTCAG-----CTCTGGGCGCCCTCGTAC 1183
 DB 520 oGlyArgProGlyGluProGlyLeuProGlyAlaAArgGlyLeuThrGlyArgProGlyAs 540
 QY 1182 AGCC-----CCTCG 1174
 DB 540 pAlaGlyProGlnGlyLysValGlyProSerGlyAlaProGlyGluAspGlyArgProG 560
 QY 1173 CCCACGAAATCCGTGTAACACAGCGTGAAGGTGCATGAGTG-----CCATC 1129
 DB 560 yProGlyProGlnGlyAlaArgGlyGlnProGlyValMetGlyPheProGlyProLy 580
 QY 1128 CAGCTGCACAGCTCAGCCACGACGCGC-----GCAGGGTGCAGGCGCATG 1081
 DB 580 sGlyAlaAsnGlyGluProGlyLysAlaGlyGlyLysGlyLeuAlaGlyAlaProGlyLe 600
 QY 1080 CGGCAGCACAGCTGTGTCACCGCGGAGCAGGCGCCAGGTTCGCGAAGCCACGCG 1021
 DB 600 uArgGlyLeuProGlyLysAspGlyGluThrGlyAlaAlaGlyProGlyProSerG 620
 QY 1020 GCCCGCATGACAGCAGTGGCGGCGACGAGGCGGCGACAGCCCTTCGTGCTGCTG 961
 DB 620 yProAlaGlyGluArgGlyGluGlnGlyAlaProGlyProSerGlyPhe----- 636
 QY 960 GTGGGGCCAGCGCTGCTCCTCAGCCACCA----- 930
 DB 637 -GlnGlyLeuProGlyProGlyProGlyProGlyGlyGlyGlyLysGlnGlyAspGlnG 656
 QY 929 -----GCAGTGTGGCTGCTACCGAGGTGAGGAGATGAGGGTGAGC 889
 DB 656 yIleProGlyGluAlaGlyAlaProGlyLeuValGlyProArgGlyGluArgGlyPhePr 676
 QY 888 AGGCCAAGAGGACACTCT-----CCTGGTG 862
 DB 676 oGlyGluArgGlySerProGlyAlaGlnGlyLeuGlnGlyProArgGlyLeuProGlyTh 696
 QY 861 CCCAGGT-----AGGGG-----CCAGGCGACTGTG 835
 DB 696 rProGlyThrAspGlyProLysGlyAlaAlaGlyProAspGlyProGlyAlaGlnG 716
 QY 834 TCCAGTCATGCGAGCA-----GGAGGTAGCCCC 805
 DB 716 yProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyIleAlaGlyProLy 736
 QY 804 AGGCAGCCCCAAGACTGTATCATGAAGCATACAGAGTAGGCTGCGGACAGTGGT--- 747
 DB 736 sGlyAspArgGlyAspValGlyGlyLysGlyProGlyGlyAlaProGlyLysAspGlyG 756
 QY 746 -----CCGGTCTCC-----GGAAGAGGTCA 727
 DB 756 yArgGlyLeuThrGlyProGlyProGlyProGlyProAlaGlyAlaAsnGlyGluLysG 776
 QY 726 GAGCAGGCGCTCCA---GTGGAGTGAAGCA-----CC 694
 DB 776 yGluAlaGlyProGlyProSerGlySerThrGlyAlaArgGlyAlaProGlyGluPr 796
 QY 693 TGGCCACAGAAGTCCAGCAGCCCGCCAGC-----CCAGGATGACAGTGGC 649
 DB 796 oGlyGluThrGlyProGlyProGlyProAlaGlyPheAlaGlyProGly---AlaAspG 815

QY 648 AGCTCCAGGGGCC---TGGATCCGGGCGACAGAGCCCTGCTAGCCAGCGCGCCCTTGG 592
 DB 815 yGlnProGlyAlaLysGlyAspGlnGlyGluAlaGly----- 827
 QY 591 ATGAGAAAGAGGCTCAGCAGGATGCCCAAG----- 561
 DB 828 ----GlnLysGlyAspAlaGlyAlaProGlyProGlnGlyProSerGlyAlaProGlyPr 846
 QY 560 -----ACAGTGCCCGCATGAGGCGCGCGCGCATAGGCTTCC 520
 DB 846 oGlnGlyProThrGlyValThrGlyProLysGlyAlaArgGlyAlaGlnGlyProProGl 866
 QY 519 CGCAGTGTCTACTGCTGAGCTAGGACGGGACACAGACAGGCG---CCAGCACTGGA 463
 DB 866 yAlaThrGlyPhe-----ProGlyAlaAlaGlyArgValGlyProGlyAlaAs 883
 QY 462 CCAATGCCAGCAGCATGCTCATGAACCTCTCTACCCCCACCTTCACAGCAGCAGAGGC 403
 DB 883 nGlyAsnProGlyProAlaGly-----ProProGlyProAlaGlyLysAs 898
 QY 402 GGCACATAGGTGATGCTCGCGCCAAACACACTCCAGGCCAAAGTTAGCAGGTGACC 343
 DB 898 pGlyProLysGlyValArgGlyAspSerGlyProGlyArg-----AlaGlyAspPr 916
 QY 342 AGCAAGAGCTGGGCTTCCGGTGCC---GCACAGGCGGCTCACCACAGCCCTCTGGACC 286
 DB 916 oGlyLeuGluGlyProAlaGlyAlaProGlyGlyLysGlyGluProGlyAspGlyPr 936
 QY 285 ATAGTGG-----CCAGGCGGTAGGCTCAGGGGCGCGTTCAGGCACACT--- 242
 DB 936 o-SerGlyLeuAspGlyProGlyProGlyProGlnGlyLeuAlaGlyGlnArgGlyIleValG 956
 QY 241 -----CCAGAACTGCTCTGCTCTCGCTCTGCTCTGCTC 214
 DB 956 lLeuProGlyGlnArgGlyGluArgGlyPheProGlyLeuProGlyProSerGlyGluP 976
 QY 213 CAGAAAGTGGCGGCT 154
 DB 976 roGlyLysGlnGlyAlaProGlyAlaSerGlyAsp-----ArgGlyProGly--- 992
 QY 153 ATTTCTGCCAGCCCTTTGTCGCGTCCAGCTTCTCAGCCCATGCTCAACACCTCTGCTGT 94
 DB 993 -----ProValGly-ProProGlyLeuThrGlyProAlaGlyGluProGlyArg 1008
 QY 93 GTGGGCGACCTCAGTGGGACACGCTCATCACTCAGATCTCTGCGCGA----- 46
 DB 1009 GluGlySerProGlyAlaAsp-----GlyProGlyArgAspGlyAlaAla 1024
 QY 45 -----GGCGCGGCTGTCTACCGCGAGCC 22
 DB 1025 GlyValLysGlyAspArgGlyGluThrGlyAlaLeuGlyAlaProGlyAla 1041

RESULT 17
 CA21_RANCA
 ID CA21_RANCA STANDARD; PRT; 1355 AA.
 AC O42350;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 2(I) chain precursor.
 GN COL1A2.
 OS Rana catesbeiana (Bull frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
 OX NCBI_TaxID=8400;
 RN (1)
 RP TISSUE=Tail;
 RC SEQUENCE FROM N.A.
 RX MEDLINE=97417499; PubMed=9272872;
 RA Asahina K., Oofusa K., Obara M., Yoshizato K.;
 RT "Cloning and characterization of the full length cDNA encoding alpha2
 RT type I collagen of bullfrog Rana catesbeiana.";
 RL Gene 194:283-289(1997).

CC CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC CC (FIBRILLAR FORMING COLLAGEN).
 CC CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 CC CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 CC CC HYDROXYAPATITE.
 CC CC -1- PTM: PROLINES ARE AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC CC
 CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).
 CC CC
 CC CC EMBL: D88764; BAA2380.1; --
 CC CC InterPro: IPR000087; Collagen.
 CC CC InterPro: IPR000885; Fib_collagen_C.
 CC CC Pfam: PF01391; Collagen; 18.
 CC CC Pfam: PF01410; COLFI; 1.
 CC CC ProDom: PD000007; Collagen; 3.
 CC CC ProDom: PD002078; Fib_collagen_C; 1.
 CC CC SMART: SM00038; COLFI; 1.
 CC CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC CC Glycoprotein; Collagen; Signal.
 CC CC SIGNAL 1 22 POTENTIAL.
 CC CC PROPEP 23 82 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
 CC CC CHAIN 83 1093 COLLAGEN ALPHA 2(I) CHAIN.
 CC CC PROPEP 1094 1355 CARBOXYL-TERMINAL PROPEPTIDE
 CC CC (BY SIMILARITY).
 CC CC CARBOHYD 1206 1206 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CC CARBOHYD 1256 1256 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CC SEQUENCE 1355 AA; 127643 MW; CB793AD5D6F4ID2A CRC64;
 CC CC
 CC CC Alignment Scores:
 CC CC Pred. No.: 6.14e-08 Length: 1355
 CC CC Score: 324.50 Matches: 304
 CC CC Percent Similarity: 32.93% Conservative: 76
 CC CC Best Local Similarity: 26.34% Mismatches: 435
 CC CC Query Match: 5.21% Indels: 342
 CC CC DB: 1 Gaps: 68
 CC CC
 CC CC US-09-759-143-110 (1-3410) x CA21_RANCA (1-1355)
 CC CC
 CC CC QY 3029 GGGAAAGTTGGGGTAGGGGAAAGTTGGGGTAGGGGAAATTTTGGGCAGTGCCTTCATC 2970
 CC CC DB 226 GlySerAspGlySerGlyProValGlyProAlaGlyProAlaGlyProIleGlySerAlaGly-- 244
 CC CC QY 2699 AGCCAGTCTTAGAGAGAGTAGAGGGAGT-----GGAGTGGGGGGAACCCAG 2922
 CC CC DB 245 AlaProGlyLeuProGlyAlaProGlyAlaLysGlyGluLeuGlyProAlaGlyAsnAsn 264
 CC CC QY 2921 GCTGGCCCAAGAGAAGAGGGGTGTAGGGAAGCCG-----TTGAGACCT 2877
 CC CC DB 265 GlyProThrGlyAlaAlaGlyGlyArgGlyGluProGlyProGlySerLeuGlyPro 284
 CC CC QY 2876 GAAGCCCAACCCCTTACCTTCCTCAACACCCCTAACCTTGGGTAAACAGCATTTTGGAAATTA 2817
 CC CC DB 285 AlaGlyProProGly-----AsnProGlyThrAsnGlyValAsnGly 298
 CC CC QY 2816 TCATTTGGGATGAGTAGAATTTCCAAAGTCTCTGGGTGA-----GGCATTTGGGGGGCCAG 2761
 CC CC DB 299 AlaLysGlyThrAlaGlyLeuProGlyValGlyGlyAlaProGlyLeuProGlyGlyArg 318
 CC CC QY 2760 ---ACCCAGGAGAGAAGATTCTGGCAATCATCAGCCCAATGA----- 2720
 CC CC DB 319 GlyIleProGly-----ProAlaGlyProAlaGlyProSerGlyAlaArgGlyLeuAla 336
 CC CC QY 2719 -----CCAGCTATCTCAGGGGACCTGATTGTGGGGATCCCCCAACCCCTACCCAAATATT 2666
 CC CC DB 337 GlyAspProGlyIleAlaGlyGly-----LysGlyAsp----- 347

QY 2665 AGACACCAACACAGAAAAAGCTAGCAATGGATTCCCTTACTTCTTACTTTTAAATAAAGTT 2606
 DB 348 -----ThrGlySerLysGlyGluProGlySer----- 356
 QY 2605 AAATATTAAATGCTGTCTCTGTGATGGCAACAGAGGACCAACAGGCGCACATCCTG 2546
 DB 357 -----ValGlyGlnGlnGlyProAlaGlyProSerGly 367
 QY 2545 ATAAAGGTAAGAGGGGTGGATCAGCAAAAAGACAGTGTCTGTGGCTGAGGGGACCTG 2486
 DB 368 GluGluGlyLysArgGlyProAsnGlyGluAlaGlySerSer----- 381
 QY 2485 GTTCTGTGTGTGTCCTCAGGACTCTCCCTACAAATAGTATGTTCAATCC 2426
 DB 382 -----GlyProSerGly----- 385
 QY 2425 ATGGAGGAGTGTTCATCTCAGAAACTCCCATGCAAGAGCTACATTAAACGAAGTGCAG 2366
 DB 385 ----- 385
 QY 2365 GTTAAGGGCTTAGAGATGGGAAACAGCTGACTGAGTTTATTACGCTCCCAAAAACCT 2306
 DB 386 AsnAlaGlyIleArg--GlyValProGlyThrArgGlyLeuProGlyProAspGlyArg 404
 QY 2305 TCTTAGTGTGTCTCACTAGGAGGCTAGCTGTAAACCTGAGCTGGTGAATCCACT 2246
 DB 405 AlaGlyGlyIleGlyProAlaGlySer-----ArgGlySerSerGlyProGly 421
 QY 2245 CGAGAGTCCCGCCATTCAGTCAGTCAGTGGAGCCCTTCT----- 2210
 DB 422 AlaArgGlyProAsnGlyAspAlaGlyArgProGlyGluProGlyLeuLeuGlyAlaArg 441
 QY 2209 GGCCTCCCTGTATTAAGTCCAGACTGAAACCCCTTTGGAAGGCTCCAGTCCAGGACCCCT 2150
 DB 442 GlyLeuPro-----GlyPhe-----SerGlySerAsn 450
 QY 2149 AGAGACTGGGGAGAGAGAGAGGGAGCGCCCGCCAGCTGTGCAGCTACGACCTCA 2090
 DB 451 GlyProGlnGly--LysGluGlyProAlaGlyProGlnGlyIleGluGlyArgSerGly 469
 QY 2089 GCAGCA-----CAGGGTGGCAGCAGAGAGCCACATTTACTTTGGCAGCAACAACTGG 2036
 DB 470 AlaAlaGlyProAlaGlyAlaArgGlyGluProGlyAlaIleGlyPheProGlyProLys 489
 QY 2035 CGGCCAGCCCGCAGCCCTACAGAGCGGGAGCTGGGACCCAGTGGACCCAGTACGCA 1976
 DB 490 GlyProAsnGlyGluProGlyLysAsnGlyAspLysGlyAsnGlnGlyProSerGlyAsn 509
 QY 1975 GGCCTCCACCCCAATGTCTGGAAGCTTTTCTACGCTGAGTATTGGCCCAAGTCGCTCTT 1916
 DB 510 ArgGlyAlaProGlyProAspGlyAsn----- 518
 QY 1915 GTCAAACTACTCTGTGTAGCAAAAGTAATGGCGACCAAGCCAGCCCTCGCGCAGACAC 1856
 DB 519 -----AsnGlyAlaGln-----GlyProAla 525
 QY 1855 CATATAGGAGTACAGACTGCTGAGCTGGACATGGAGCCCATAAACAGGATG---- 1800
 DB 526 GlyLeuGlyGly--AlaThrGlyGluGlyGlnGlnGlyPro--SerGlyAlaProG 544
 QY 1799 -----GGGCCACTGGGACAGCAGCAGGAGGACCTATCCAG 1766
 DB 544 LysPheGlnGlyLeuProGlyProGlyProGlyGluValGlyLys-----ProG 562
 QY 1765 GATGGGAGGTCCAGCAGATGCCCGGCCG---GAACACCCCTGGCTCGGTGGGCTC 1709
 DB 562 LysGluArgGly-----AlaProGlyAspPheGlyProPro--GlySerAlaGly-- 577
 QY 1708 ACCCACCACACAGTACGAGAGATCATCAGGAGGAGGCGCCCGCAGGCGCGGTGAGG 1649
 DB 578 -----ThrArgGlyGluArgGly-----AlaProGlyGluSerGlyAl 591

RP SEQUENCE FROM N.A.
RX MEDLINE-90067946; PubMed=2587287;
RT Su M.W., Lee B., Ramirez F., Machado M., Horton W.;
RT "Nucleotide sequence of the full length cDNA encoding for human type
RT II procollagen.";
RL Nucleic Acids Res. 17:9473-9473(1989).
RN [2]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE-87031574; PubMed=3021582;
RT Nunez A.M., Kohno K., Martin G.R., Yamada Y.;
RT "Promoter region of the human pro-alpha 1(II)-collagen gene.";
RL Gene 44:111-116(1986).
RN [3]
RP SEQUENCE OF 432-1145 FROM N.A.
RX Ramirez F.;
RL Submitted (DEC-1988) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 963-1418 FROM N.A.
RX MEDLINE-85190534; PubMed=3857598;
RT Cheah K.S.E., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;
RT "Identification and characterization of the human type II collagen
RT gene (COL2A1).";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
RN [5]
RP SEQUENCE OF 1120-1398 FROM N.A.
RX MEDLINE-85306861; PubMed=3840017;
RT Elima K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J.,
RT Vuorio E.;
RT "Construction and identification of a cDNA clone for human type II
RT procollagen mRNA";
RL Biochem. J. 229:183-188(1985).
RN [6]
RP SEQUENCE OF 1106-1418 FROM N.A.
RX MEDLINE-88067771; PubMed=2825137;
RT Elima K., Vuorio T., Vuorio E.;
RT "Determination of the single polyadenylation site of the human pro
RT alpha 1(II) collagen gene";
RL Nucleic Acids Res. 15:9499-9504(1987).
RN [7]
RP SEQUENCE OF 1227-1289 FROM N.A.
RX MEDLINE-86104139; PubMed=3002437;
RT Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
RT "Isolation and partial characterization of genomic clones coding for
RT a human pro-alpha 1 (II) collagen chain and demonstration of
RT restriction fragment length polymorphism at the 3' end of the gene.";
RL Biochemistry 24:6343-6348(1985).
RN [8]
RP SEQUENCE OF 1176-1226 FROM N.A.
RX MEDLINE-84118798; PubMed=6320112;
RT Strom C.M., Upholt W.B.;
RT "Isolation and characterization of genomic clones corresponding to
RT the human type II procollagen gene.";
RL Nucleic Acids Res. 12:1025-1038(1984).
RN [9]
RP SEQUENCE OF 35-167 FROM N.A.
RX MEDLINE-89233138; PubMed=2714801;
RT Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
RT "Organization of the exons coding for pro alpha 1(II) collagen N-
RT propeptide confirms a distinct evolutionary history of this domain of
RT the fibrillar collagen genes.";
RL Genomics 4:438-441(1989).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE-91184577; PubMed=2010058;
RT Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE-97255959; PubMed=9101290;
RT Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-

RT associated collagen (type IX), and network-forming collagen (type X),
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RP VARIANT SER-1074.
RX MEDLINE-90036909; PubMed=2572591;
RT Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
RT Hollister D.W.;
RT "Glycine to serine substitution in the triple helical domain of pro-
RT alpha 1 (II) collagen results in a lethal perinatal form of short-
RT limbed dwarfism.";
RL J. Biol. Chem. 264:18265-18267(1989).
RN [13]
RP VARIANT SEDC 1095-GLY--TYR-1330 DEL.
RX MEDLINE-89266907; PubMed=2543071;
RT Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;
RT "Identification of the molecular defect in a family with
RT spondyloepiphyseal dysplasia.";
RL Science 244:978-980(1989).
RN [14]
RP VARIANT OSTEOARTHRTIS CYS-650.
RX MEDLINE-90370826; PubMed=1975693;
RT Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
RT "Single base mutation in the type II procollagen gene (COL2A1) as a
RT cause of primary osteoarthritis associated with a mild
RT chondrodysplasia.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
RN [15]
RP VARIANT OI-IV VAL-717.
RX MEDLINE-91291136; PubMed=2064612;
RT Bateman J.F., Hannagan M., Chan D., Cole W.G.;
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT valine substitution in osteogenesis imperfecta type IV. Detection of
RT the mutation and prenatal diagnosis by a chemical cleavage method.";
RL Biochem. J. 276:765-770(1991).
RN [16]
RP VARIANT OSTEOARTHRTIS CYS-650.
RX MEDLINE-91086471; PubMed=1985108;
RT Eyre D.R., Weis M.A., Moskowitz R.W.;
RT "Cartilage expression of a type II collagen mutation in an inherited
RT form of osteoarthritis associated with a mild chondrodysplasia.";
RL J. Clin. Invest. 87:357-361(1991).
RN [17]
RP VARIANT HYPOCHONDROGENESIS GLU-984.
RX MEDLINE-93054548; PubMed=1429602;
RT Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
RT Cohn D.H., Eyre D.R.;
RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
RT 1(II) chain produces hypochondrogenesis.";
RL J. Biol. Chem. 267:22522-22526(1992).
RN [18]
RP VARIANT HYPOCHONDROGENESIS SER-705.
RX MEDLINE-92262484; PubMed=1374906;
RT Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
RT Ramirez F., Vitale E., Lee B.;
RT "Characterization of a type II collagen gene (COL2A1) mutation
RT identified in cultured chondrocytes from human hypochondrogenesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
RN [19]
RP VARIANT WS-II ASP-198.
RX MEDLINE-93304428; PubMed=8317498;
RT Koerkoe J., Ritvaniemi P., Haataja L., Kaeaelaeninen H.,
RT Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
RT for glycine alpha 1-67 and that causes cataracts and retinal
RT detachment: evidence for molecular heterogeneity in the Wagner
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
RL Am. J. Hum. Genet. 53:55-61(1993).
RN [20]
RP VARIANT SEMD CYS-840.
RX Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,
RT Eyre D.R.;
RT "A dominant mutation in the type II collagen gene (COL2A1) produces

Db 417 yGluGluGlyLysArgGlyAlaArgGlyGluProGlyGlyValGlyProGlyLeuGlyProGlyPro 437
 QY 906 AGGAAGATGAGGTGAGCAGGCAAGAGGC-----ACTCC 871
 Db 437 oGlyGluArgGlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaGlyPro 457
 QY 870 TCCTGGTCCCGAGT---AGGGGCCAGGCGACTGTGTCCAGTCAATGGCAGGCGAG 814
 Db 457 oLysGlyAlaProGlyGluArgGlyProSerGlyLeuAlaGlyProLysGlyAlaAsnGly 477
 QY 813 A-----GGTAGCCAGCAGGCCCAAGACTGATCATGAAGGCATAGACAGATAGGCC 760
 Db 477 yAspProGlyArgProGlyGluProGlyLeuProGlyAlaArgGlyLeuThrGlyArgPro 497
 QY 759 TGCGCAGACTGTCCGGTCCCGGAAGAGGTGAGAGCAGGCGCTCCAGTGGAGTGAAG 700
 Db 497 oGlyAspAlaGlyPro---GlnGlyLysValGlyProSerGlyAlaProGlyGluAspGly 516
 QY 699 CACACTGGCCAGAGAAGTCCAGAGCCCGCCAGC-----CC 664
 Db 516 yArgProGlyProProGlyGlnGlyAlaArgGlyGlnProGlyValMetGlyPhePro 536
 QY 663 AGATGAGAGTCCAGCTCCAGGCGCTGGATCGGCGCAGCAGCAGCCCTGCTAGCCAG 604
 Db 536 oGly-----ProGlyGlyAlaAsnGlyGluProGlyLysAlaGly----- 549
 QY 603 CCGGCCCTTGGGATGAGAAGAGGCTTCAGCAGGATGCCA-----AGGACAGTCCCGCAG 550
 Db 550 -----GluLysGlyLeuProGlyAlaProGlyLeuArgGlyLeuPro----- 563
 QY 549 ATGAAGGCGCGGCGGCGCCATAGCTCCAGCCAGTGTCTACTGGCTCAGCTAGGAGC 490
 Db 564 -----GlyLysAspGlyGluThrGlyAla 571
 QY 489 GGGACACAGCAGGCGCCAGCA-----CTGGACCAATGCCCGCAGC 451
 Db 571 aGluGlyProProGlyProAlaGlyProAlaGlyGluArgGlyGluGlnGlyAlaProGly 591
 QY 450 ACCATGTGTATGAACCTTCTCTCTACCCCACTTCAGCAGCAGAGGCGGCACATAGTGT 391
 Db 591 yPro-----SerGlyPheGlnGlyLeuProGly----- 600
 QY 390 ATGCTCGCGCCAAACACACTCCAGGCCAAAGTTAGCAGGTTGACCAAGCAAGAGCTGG 331
 Db 601 -----ProProGlyProProGlyGlu-----GlyGlyLysProGlyAspGlnGly 615
 QY 330 GCTTTCCGGTCCCGCAGCAGGC----- 309
 Db 615 yValProGlyGluAlaGlyAlaProGlyLeuValGlyProArgGlyGluArgGlyPhePro 635
 QY 308 -----GGCTACCCACAGCCTCTGA-----CCATAGTG 280
 Db 635 oGlyGluArgGlySerProGlyAlaGlnGlyLeuGlnGlyProArgGlyLeuPro-GlyT 655
 QY 279 GCGCAGGCGGTAGGCGGTTCAGGGGCGGTTCAGGCACCTCCAGAACTGCTGCTCGGCT 220
 Db 655 hrProGlyThrAspGlyProLysGlyAlaSerGlyPro-----A 668
 QY 219 CTGCTCAGAGCTCGCGCCCTCTCTCTCTGCTCCGCCCACTGCTAGGAATCAGCCAG 160
 Db 668 laGlyProProGlyAlaGlnGlyProProGlyLeuGlnGlyMetProGlyGluArgGlyAla 688
 QY 159 GCGCCATTTCTGCAGCAGCCTTTGCT-----GCCGTCACGCTT 121
 Db 688 laAlaGlyLeuAlaGlyProLysGlyAspArgGlyAspValGlyGluLysProGluGlu 708
 QY 120 CTCAGCCC-----ATGCTCAACACTGCTGCTGTG 91
 Db 708 lyAlaProGlyLysAspGlyArgGlyLeuThrGlyProGlyProGlyPro 727
 QY 90 GGGCACCCTCAGTGGGAGCACCTCTCATCTAGCTCCTGCT-----CGAGGCG 43

QY 2288 ACTAGGAGCTAGCTGTTAAACCTTGAGCTGGGTAAATCACTGACAGTCCCGCATTC 2229
Db 435 -----GlySerProGlyThrSerGlyProGly 444
QY 2228 CAGTGCATGAGCCCTTGGCTCCTGTATAAGTCCAGACTGAACCCCTTGAAGG 2169
Db 444 ySerAla---GlyProProGlySerProGly---ProGlnGlySerThrGlyProGlnG 462
QY 2168 CTCCAGTCAGGAGCCCTAGAGACTGGGAGAGAGG-----AGAGGGAC 2124
Db 462 y---AsnSerGlyLeuProGlyAspProGlyPheLysGlyGluAlaGlyProLysGlyG 481
QY 2123 GCCCCAGCCCGCTGCTGAGCTAGCAGCTCAGCAGCAGCAGGCTGGCAGCAGAGCC 2064
Db 481 uProGlyProHisGlyLeuGln----- 488
QY 2063 ACATTACTTTGGCAGCAACAGAACTGGCGGCAGCCCGGCAGCCCATGGGCTAACAG 2004
Db 489 -----GlyProLleGlyProProGlyGluGluGly 499
QY 2003 GAGCGGGAGCTGGGA---CCAGTGAGCAGGCCCTCCACCCCAATGTCTGGAAGTTT 1947
Db 499 sArgGlyProArgGlyAspProGlyThrLeuGlyProPro----- 512
QY 1946 TCTACGCTGAGTATTTGGCGAAGTCTCTTGTCAAAATACCTACCTGTGTAGCAAAAGTAA 1887
Db 513 -----GlyProVal----- 515
QY 1886 TGCGCAGCAGCCAGCCCTGGCGCAGACACCATATAGCAGTGCAGACACTG----- 1835
Db 516 -GlyGluArgGlyAlaProGlyAsnArgGlyPheProGlySerAspGlyLeuProGlyPr 535
QY 1834 -----GCTGAGCTGGCAATGGAGCCCATAAACAGGATGGGCCCTGGGACAGCA 1782
Db 535 oLysGlyAlaGlnGlyGluArgGly-ProValGly-SerSerGlyProLysGlySerGln 554
QY 1781 GGAAGCAGCTATCCAGATGGGAGGTCCAGCAGATGCCCGCGCCGGAACCCCTGG 1722
Db 555 Gly-----AspProGlyArgProGlyGluPro-GlyLeuProGly 567
QY 1721 CTTGGTGGCTCACCACACACACACGTCAGGAGACATCAGAGGAGCA---GGCCCGGC 1665
Db 567 yAlaArgGlyLeuThrGlyAsnProGlyValGlnGlyProGlyLysLeuGlyProLe 587
QY 1664 AGAGCGGGGTGAGGTGGAGAGCCACTGCTCCAGCAGCCACGCTGTCCATTAGG- 1606
Db 587 u-GlyAlaProGlyGluAspGlyArgProGlyProGly-----SerIleGlyI 604
QY 1605 --AAGGAGCTCCAGGC---TTAGG---CCTGGCAGGAGCTGTGTCATCAGCTGTCT 1554
Db 604 leLysGlyProGlyThrMetGlyLeuProGlyProLys-GlySerAsnGlyAspPro 623
QY 1553 CACTGTAGCAGCTCCAGTGTCCCTCGGTATT-----TGGCAGGAAACCTGCTTCT 1500
Db 624 -----GlyLysProGlyGluAlaGlyAsnProGlyVal 634
QY 1499 CCGGTGGTAGAGGAGG-----CCAGTGTGTAGGAGCA 1464
Db 635 ProGlyGlnArgGlyAlaProGlyLysAspGlyLysValGlyProGlyProGly 654
QY 1463 TCTGCAAGGCTGAGAGGTGACCCCGTGGCGGCTGAAGCTGTCCAGCCAGCCACAC 1404
Db 655 ProProGlyLeuArg-GlyGluArgGlyGluGlnGly---ProProGlyProThrGlyPhe 673
QY 1403 TGTGGCAGCAGGATGTGGCAGCGGCAGCCAGCAGGAGAAAG----- 1365
Db 674 GlnGly-----HisProGlyProProGlyProProGlyGluGlyLysPro 689
QY 1364 -----CTGCCACACTGGCCAAATAGACTGCTCGAAGTGGCAATGCTGCACCA 1317
Db 690 GlyAspGlnGlyValProGlyGlyProGlyAlaVal-----GlyPro 703
QY 1316 GCCGGTCCATGACCAGAGAGA---AGACCAGGAGATGGCGACTGCAGGAACACCCCA 1260

Db 704 LeuGlyProArgGlyGluArgGlyAsnProGlyGluArgGlyGluProGlyIleThr--- 722
QY 1259 GGCTGCCATCCGAACGC-----CTTCATCATAGTGTCTCCGGCTCGG----- 1215
Db 723 GlyLeuProGlyGluLysGlyMetAlaGlyGlyHisGlyProAspGlyProLysGlySer 742
QY 1214 -----TGCCCGGCTCAGCTCTGGGACGCGCTGGT 1185
Db 743 ProGlyProSerGlyThrProGlyAspThrGlyProGlyLeuGlnGlyMetProGly 762
QY 1184 ACA-----GCCCTCGCCAGCAAAATCGGTGTAACACAGCGTGAAGTCA 1140
Db 763 GluArgGlyIleAlaGlyThrProGlyProLysGlyAsp----- 775
QY 1139 TGAGTGCCATCCAGCTGCACAGCTCAGCCAGGAAGAGCGCGCAGGTCGGGGCATGC 1080
Db 776 -----ArgGlyGlyIleGlyGluLysGlyAla 784
QY 1079 GGCAGCAGCTGTGTGTCAGCCGGGGAAGCAGGCGCCAGGTTCGGAAGCCAGCGG 1020
Db 785 GluGlyThrAlaGlyAsnAspGlyAlaGlyGlyLeuProGly-----ProLeuGly 801
QY 1019 CCCGCATGGACAGCAGTGGGGCCACAGGAGGGGGCGGACAGCCCTTCTGCTGGCTCGG 960
Db 802 Pro-----ProGlyProAlaGlyLeu----- 809
QY 959 TGGGCCCAGCGCTGCCTCCTCAGCCAGCAGCTGTGTGCTACGAGGTGAGGAAGA 900
Db 810 --GlyGluLysGlyGluProGlyProArg-----GlyLeu 820
QY 899 TGAGGTGAGCAGGCCAAAGAGGACACTCCT----- 870
Db 821 ValGlyProProGlySerArgGlyAsnProGlySerArgGlyGluAsnGlyProThrGly 840
QY 869 -----CCTGGTGCCAGGTAGGGGG 849
Db 841 AlaValGlyPheAlaGlyProGlnGlySerAspGlyGlnProGlyValLysGlyGlu--- 859
QY 848 CCAGGGCACTGTGTCCAGTATGGCAGGAGGTAGCCAGCAGCCAGCCCAAGAC 789
Db 860 ProGlyGlu-----ProGlyGlnLysGlyAspAlaGlySerProGly---ProGlnGly 876
QY 788 TGATCATGAAGGCATAGACAGAGTAGGCTGGCCAGCAGTGTGTCTCGGCT----- 741
Db 877 -----LeuAlaGlySerProGlyProHisGlyProAsnGlyValProGly 891
QY 740 -----CCCGGAAGAGT 729
Db 892 LeuLysGlyGlyArgGlyThrGlnGlyProProGlyAlaThrGlyPheProGlySerAla 911
QY 728 CAGAGCAGGCGCTCCAGTGGAGTGAGCAGCACCTGGCCACAGAAAGTCCAGCAGCCCA 669
Db 912 GlyArgValGlyProProGlyProAlaGlyAlaProGly-----ProAlaGlyPro 928
QY 668 -----GCCCGCAGATGAGCAGTCCAGCTCCAGGCGCTGGGATCCGGGACAGCAGCC 615
Db 929 LeuGlyGluProGlyLysGluGlyProProGlyProArgGlyAspProGlySerHisGly 948
QY 614 CTGTAGCAGCCCGCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTG 555
Db 949 ArgValGly-----ValArgGlyProAlaGlyProProGly----- 960
QY 554 CCCAGATGAAGGGCGCGCGGCATAGCTGCCAGCCAGTGTGTCTAGGCTAGCCTA 495
Db 960 ----- 960
QY 494 GGAGGGGACACAGCAGCCAGCAGCTGGACCAATGCCAGCAGCAGCATGTGTGAAGT 435
Db 961 GlyProGlyAspLysGlyAspProGlyGluAspGlyGlnProGlyProAspGly----- 978
QY 434 TCTCTCTACCCCTTCCAGCAGCAGGCGGCACATAGTGTATGCCTCGGCGCAAC 375

Db 979 -----ProGlyProAlaGlyThrThrGlyGlnArgGlyIleVal----- 992
 QY 374 ACACCTCAGCCCAAGG-----TTAGCAGGTGA 345
 Db 993 GlyMetProGlyGlnArgGlyGluArgGlyMetProGlyLeuProGlyProAlaGlyThr 1012
 QY 344 CCAGCAGAGCTGGGCTTCGCGTCCGCGCAGCAGGC-----GGC 306
 Db 1013 ProGlyGlyValGlyProThrGlyAlaThrGlyAspGlyGlyProGlyProValGly 1032
 QY 305 TCACCCACAGCCTCGACCATAGTGGG---CCAGCGGGTAGGGCTCAGGGGGCCGCTTC 249
 Db 1033 ProGlySerAsnGlyPro-ValGlyGluProGlyProGluGlyProAlaGlyAsnAs 1052
 QY 248 AGGCACCTCCAGAACTGCTGCTCGCTGCTGCTCCAGAAAGCTCGCGCTCTCTCTCTCTG 189
 Db 1052 pGlyThrProGly-----ArgaspGlyAlaValGlyGluArgGly----- 1065
 QY 188 CTGCGCGCACTGCTAGGAATCAGCCAGCGCCCATTTCTGCGAGCCCTTTGTCGCGG 129
 Db 1066 -----AspArgGlyAspProGly-----ProAlaGlyLeuPr 1076
 QY 128 TCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGCGACCTCAGTGGGGACACGT 69
 Db 1076 oGlySerGln-GlyAlaProGlyThrProGlyProValGly---AlaProGlyAspAlaG 1095
 QY 68 CTCATCACTCAGATCCTGCGC-----CGAGCGCGCG 39
 Db 1095 lyGlnArgGlyAspProGlySerArgGlyProIleGlyHisLeuGlyArgAlaGlyLysA 1115
 QY 38 GGCTGTACCCGCGCAGCGCCGCTGCAGG 10
 Db 1115 rgGlyLeuProGlyProGlnGlyProArg 1124

RESULT 20
 CA54 HUMAN
 ID CA54_HUMAN STANDARD; PRT: 1585 AA.
 AC P29400; Q16126; Q16006;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 5(IV) chain precursor.
 GN COL4A5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94165049; PubMed=8120014;
 RA Zhou J., Leinonen A., Tryggvason K.;
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen
 chain and identification of a single-base mutation in exon 23
 converting glycine 521 in the collagenous domain to cysteine in an
 Alport syndrome patient.";
 RL J. Biol. Chem. 269:6608-6614(1994).
 RN [2]
 RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
 RC TISSUE=Kidney;
 RX MEDLINE=92316923; PubMed=1352287;
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen
 chain and identification of a single-base mutation in exon 23
 converting glycine 521 in the collagenous domain to cysteine in an
 Alport syndrome patient.";
 RL J. Biol. Chem. 267:12475-12481(1992).
 RN [3]
 RP SEQUENCE OF 85-1685 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=90337990; PubMed=2380186;
 RA Pihlajaniemi T., Pohjola-Erila E., Myers J.C.;
 RT "Complete primary structure of the triple-helical region and the
 carboxyl-terminal domain of a new type IV collagen chain, alpha
 5(IV).";
 RL J. Biol. Chem. 265:13758-13766(1990).
 RN [4]

RP SEQUENCE OF 924-1685 FROM N.A.
 RX MEDLINE=91169491; PubMed=2004755;
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
 RT "Characterization of the 3' half of the human type IV collagen alpha
 5 gene that is affected in the Alport syndrome.";
 RL Genomics 9:1-9(1991).
 RN [5]
 RP SEQUENCE OF 914-1685 FROM N.A.
 RX MEDLINE=90160375; PubMed=1689491;
 RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,
 Tryggvason K.;
 RT "Identification of a distinct type IV collagen alpha chain with
 restricted kidney distribution and assignment of its gene to the
 locus of X chromosome-linked Alport syndrome.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
 RN [6]
 RP SEQUENCE OF 1442-1471 FROM N.A.
 RX MEDLINE=90252791; PubMed=2339699;
 RA Myers J.C., Jones T.A., Pohjola-Erila E., Kadri A.S., Goddard A.D.,
 Sheer D., Solomon E., Pihlajaniemi T.;
 RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
 to the region of the X chromosome containing the Alport syndrome
 locus.";
 RL Am. J. Hum. Genet. 46:1024-1033(1990).
 RN [7]
 RP SEQUENCE OF 1-20 FROM N.A.
 RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,
 Marynen P.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).
 RX MEDLINE=94133540; PubMed=8301933;
 RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
 Cassiman J.-J., Marynen P.;
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood
 cells: a complex mutation in the COL4A5 gene of an Alport patient
 deletes the NC1 domain.";
 RL Kidney Int. 44:1316-1321(1993).
 RN [9]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97338662; PubMed=9195222;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 RT "The clinical spectrum of type IV collagen mutations.";
 RL Hum. Mutat. 9:477-499(1997).
 RN [10]
 RP VARIANT AS SER-1564.
 RX MEDLINE=91169492; PubMed=1672282;
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,
 Tryggvason K.;
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a
 conserved cysteine to serine in Alport syndrome.";
 RL Genomics 9:10-18(1991).
 RN [11]
 RP VARIANT AS ARG-325.
 RX MEDLINE=92303559; PubMed=1376965;
 RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,
 Tryggvason K., Gubler M.-C., Antignac C.;
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5
 (IV) chain associated with X-linked Alport syndrome: characterization
 of the mutation by direct sequencing of PCR-amplified lymphoblast
 cDNA fragments.";
 RL Am. J. Hum. Genet. 51:135-142(1992).
 RN [12]
 RP VARIANT AS GLU-325.
 RX MEDLINE=93244772; PubMed=1363780;
 RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,
 Rizzoni G.F., de Marchi M.;
 RT "De novo mutation in the COL4A5 gene converting glycine 325 to
 glutamic acid in Alport syndrome.";
 RL Hum. Mol. Genet. 1:127-129(1992).
 RN [13]
 RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
 RX MEDLINE=94010948; PubMed=8406498;

RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
 RA Tryggvason K., Hagsma-Schouten W.A.G., Roodvoets A.P., Rascher W.,
 van Oost B.A., Smeets H.J.M.;
 RT "Identification of four novel mutations in the COL4A5 gene of
 RT patients with Alport syndrome.";
 RL Genomics 17:485-489(1993).
 RN [14]
 RP VARIANTS AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872
 RP AND C-1241.
 RA MEDLINE-95322976; PubMed-7599631;
 RX Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
 RT "Detection of 12 novel mutations in the collagenous domain of the
 RT COL4A5 gene in Alport syndrome patients.";
 RL Hum. Mutat. 5:197-204(1995).
 RN [15]
 RP VARIANT AS ARG-1649.
 RX MEDLINE-96213750; PubMed-8651292;
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
 Denison J.C., Fain P.R., Gregory M.C.;
 RT "A mutation causing Alport syndrome with tardive hearing loss is
 RT common in the western United States.";
 RL Am. J. Hum. Genet. 58:1157-1165(1996).
 RN [16]
 RP VARIANTS AS.
 RX MEDLINE-96213754; PubMed-8651296;
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
 Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
 Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
 Savi M., Ballabio A., de Marchi M.;
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
 RT exons of the COL4A5 gene.";
 RL Am. J. Hum. Genet. 58:1192-1204(1996).
 RN [17]
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
 RP MET-1428
 RX MEDLINE-97094179; PubMed-8940267;
 RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,
 Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broeyer M.,
 Gubler M.-C., Antignac C.;
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
 RT syndrome.";
 RL Am. J. Hum. Genet. 59:1221-1232(1996).
 RN [18]
 RP VARIANT AS ASP-1498.
 RX MEDLINE-96233932; PubMed-8829632;
 RA Tverskaya S., Bobrylina V., Tsalykova F., Ignatova M.,
 Krasnopolskaya X., Evgrafov O.;
 RT "Substitution of A1498D in noncollagen domain of a5(IV) collagen
 RT chain associated with adult-onset X-linked Alport syndrome.";
 RL Hum. Mutat. 7:149-150(1996).
 RN [19]
 RP VARIANT AS GLN-1677.
 RX MEDLINE-97295089; PubMed-9150741;
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
 RT "Common ancestry of three Ashkenazi-American families with Alport
 RT syndrome and COL4A5 R1670.";
 RL Hum. Genet. 99:681-684(1997).
 RN [20]
 RP VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.
 RX MEDLINE-98112435; PubMed-9452056;
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
 Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
 Trivelli A., Pinciaroli A.R., Ragaio M., Rizzoni G.F., de Marchi M.;
 RT "Missense mutations in the COL4A5 gene in patients with X-linked
 RT Alport syndrome.";
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).
 RN [21]
 RP VARIANTS AS V-420; 456-P-P-458 DEL; D-573; D-624; D-635; 802-G--P-807
 RP DEL; R-869; C-941; S-1030; S-1066; D-1143; R-1196; E-1261; S-1357
 RP AND R-1649.
 RX MEDLINE-99063529; PubMed-9848783;
 RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
 Barker D.F., Gregory M.C., Atkin C.L., Styrkarsdottir U., Neumann H.,

RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected
 RT Alport syndrome using PCR and direct DNA sequencing.";
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
 RN [22]
 RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
 RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
 RX MEDLINE-20030197; PubMed-10561141;
 RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
 Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
 RT "Detection of mutations in the COL4A5 gene in over 90% of male
 RT patients with X-linked Alport's syndrome by RT-PCR and direct
 RT sequencing.";
 RL Am. J. Kidney Dis. 34:854-862(1999).
 RN [23]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [24]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [25]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [26]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [27]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [28]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [29]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [30]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [31]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [32]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [33]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [34]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [35]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [36]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [37]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [38]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [39]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [40]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [41]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [42]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [43]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [44]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [45]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [46]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [47]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [48]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [49]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [50]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [51]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [52]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [53]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [54]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [55]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [56]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [57]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [58]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [59]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [60]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [61]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [62]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [63]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [64]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [65]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [66]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [67]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [68]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [69]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [70]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [71]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [72]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [73]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [74]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [75]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [76]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [77]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [78]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [79]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [80]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [81]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [82]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [83]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [84]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [85]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [86]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [87]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [88]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [89]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [90]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [91]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [92]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [93]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [94]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [95]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [96]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [97]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [98]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [99]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [100]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [101]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [102]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [103]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [104]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [105]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [106]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [107]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [108]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [109]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [110]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [111]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [112]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [113]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [114]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [115]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [116]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [117]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [118]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [119]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [120]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [121]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [122]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [123]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [124]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [125]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [126]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [127]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [128]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [129]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [130]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [131]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [132]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [133]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [134]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [135]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [136]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [137]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [138]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [139]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [140]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [141]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [142]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [143]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [144]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [145]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [146]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [147]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [148]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [149]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [150]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [151]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [152]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [153]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [154]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [155]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [156]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [157]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [158]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [159]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [160]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [161]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [162]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [163]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [164]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [165]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [166]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [167]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [168]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [169]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [170]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [171]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [172]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [173]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [174]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [175]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [176]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [177]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [178]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [179]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [180]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [181]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [182]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [183]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [184]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [185]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [186]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [187]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [188]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [189]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [190]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [191]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [192]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [193]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [194]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [195]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [196]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [197]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [198]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [199]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [200]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [201]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [202]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [203]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [204]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [205]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [206]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [207]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [208]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [209]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [210]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [211]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [212]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [213]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [214]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [215]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [216]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [217]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [218]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [219]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [220]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [221]
 RP VARIANT AS ARG-822.
 RX MEDLINE-20025011; PubMed-10563487;
 RN [222]
 RP VARIANT AS ARG-822.
 RX MEDLINE-200

Db 653 lYAspProGlyGlnThrIleThrGlnPro-----GlyLysProGlyLeuP 668
 QY 656 CCGATCCAGGCCCTGGAGCTGCTCATCTCTGG-----CGTGGG 673
 Db 668 roGlyAsnProGlyArAspGlyAspValGlyLeuProGlyAspProGlyLeuProGlyG 688
 QY 674 CTGCTGGACTTCTGTGGCCAGGTGTCTTCACTCCACTGGAGGCCCTGTCTGTACCT- 732
 Db 688 lNProGlyLeu-----ProGlyIleProGlySerLysGlyGluProGlyIle-----ProG 705
 QY 733 -----CTTCCGGGA-----CCCGGACCACTGTG-----CCAGGCCTACTCTGTATGCC 778
 Db 705 lYlleglyLeuProGlyProProGlyProLysGlyPheProGlyIlePro-----721
 QY 779 TTCATGATCACTTGTGGCTGCTGGCTACTCTCTGCTGCCCTGCTGCTGCTGCTGCTG 838
 Db 722 -----GlyProGlyAlaProGlyThrProGlyArgIleGlyLeuG 736
 QY 839 AGTCCCTGGCCCTACCTGGGACCCAGGA-----GGAGTGC 877
 Db 736 luGlyProProGlyProProGlyPheProGlyProLysGlyGluProGlyPheAlaLeuP 756
 QY 878 CTCTTTGGCTGCTACCTCATCTCTCCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 937
 Db 756 roGlyProProGlyProProGlyLeuProGlyPheLysGlyAlaLeuGlyProLysGlyA 776
 QY 938 GAGGAGCAGCGCTGGGCCCCACCGACCCAGCAGAGGGCTGCTGGCCCTCTTGTCTG 997
 Db 776 sPArGlyPheProGlyPro-----ProGlyProProGlyArgT 789
 QY 998 CCCCACTGTGTCATGCCGGGCGCTTGGCTTTCGGAACTGGGCGCCCTGCTTCC 1057
 Db 789 hrGlyLeuAspGlyLeuProGlyProLysGlyAspValGlyProAsnGlyGlnProGlyP 809
 QY 1058 CGGCTGACCACTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1105
 Db 809 roMetGlyProProGlyLeuProGlyIleGlyValGlnGlyProProGlyProProGlyI 829
 QY 1106 TTGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1165
 Db 829 leProGlyProIleGlyGln-----ProGlyLeuHisGly- 840
 QY 1166 TTGCTGGGAGGGCTGTACAGCGCTGCCAGAGCTGAGCGGGCCAGCGGCCCGG 1225
 Db 841 -----IleProGlyGlu-----LysGlyAspProGlyProp 851
 QY 1226 AGACACTATGATGAGGCGTTCGGATGGCAGCGCTGGTGTCTGCTGCTGCTGCTGCTG 1285
 Db 851 roGlyLeu-----AspValProGlyPro-----ProGlyGluArgGlyS 864
 QY 1286 TCCCTGTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1345
 Db 864 erProGlyIleProGlyAlaProGlyProIleGlyProPro-----GlySerProGlyL 882
 QY 1346 TATTTGGCAGTGTGCGAGC-----TTTCCCTGTGCTGCTGCTGCTGCTGCTGCTGCT 1399
 Db 882 euProGlyLysAlaGlyAlaSerGlyPhePro-----GlyThrLysGlyGluMetGlyMet- 900
 QY 1400 CACAGTGTGGCGTGTGACAGCTTTCAGCGCCCTCACCAGGCTTACCTTCTCAGCCCTG 1459
 Db 900 -----900
 QY 1460 CAGATCCTGCCCTACACTGGCTTCCCTTCTACACCGGGAGAGCAGGTGTCTGCTGCC 1519
 Db 901 -----MetGlyProProGlyProProGlyProLysGlyIleProGlyA 915
 QY 1520 AATACCGAGGGGACACTGGAGTGTACAGTGTGAGGACAGCCTGATGACAGCTTCTG 1579
 Db 915 rg-----SerGly-ValProGlyLeuLysGlyAspGlyLeuGln-----928
 QY 1580 CCAGGCCCTTAAGCCTGGAGCTCCCTTCCCTTAATGACAGCTGGTGTGCTGGAGCAGTGC 1639

Db 929 -----GlyGlnProGlyLeuProGlyProThrGlyGluLysGlySerLysGlyGluPro 946
 QY 1640 CTCTCTCCCACTCCACCCGCGCTCTCGGGGCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1699
 Db 947 GlyLeuProGlyPro-Pro-----GlyPro-----MetAspProAsnLeuLe 960
 QY 1700 GTGGTGGGTGAGCCCAACCGAGCCAGCGGTGGT-----TCGGGCG 1738
 Db 960 u-GlySerLysGlyGluLysGlyGluProGlyLeuProGlyIleProGlyValSerGlyP 980
 QY 1739 CG---CGGCAT-----CTGCCCTGGA---CCTCGCCATCCTGGATAGTGC---CTTCC 1783
 Db 980 roLysGlyTyrlGlnGlyLeuProGlyAspProGlyGlnProGlyLeuSerGlyGlnProG 1000
 QY 1784 CTGTCCCAAGTGGCCCACTCCCTGTATGAGGCTCCATTTGCTCCAGCTCAGCCAGCTG 1843
 Db 1000 lYLeuProGlyProProGlyProLysGlyAsnProGlyLeuProGlyGlnPro-----1017
 QY 1844 ACTGCCATATGCTGCTGCCCGAGCCCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1903
 Db 1018 -----GlyLeuIleGlyProProGly-----LeuLysG 1027
 QY 1904 GTAGTATTTTACAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACAT 1963
 Db 1027 lYThr-----11eG 1030
 QY 1964 GGTGGAGGGCTGCTCACTGGTCCCGCTCCCGCTCCCTGCTGCTGCTGCTGCTGCTGCT 2023
 Db 1030 lYAspMetGlyPhePro-GlyProGlnGlyValGluGlyProProGlyProSerGlyVal 1049
 QY 2024 CCGGGTGGCCCGCACTTCTGCTGCCAAAGTAATGTGCTCTGCTGCTGCTGCTGCTGCT 2083
 Db 1050 ProGlyGlnProGlySerProGly-LeuProGly-----1060
 QY 2084 TCTGCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2143
 Db 1061 -----GlnLysGlyAspLysGlyAspProGlyIleSer-----1071
 QY 2144 GTCTCTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2200
 Db 1072 -SerIleGlyLeuProGlyLeuProGlyProLysGlyGluProGlyLeuProGlyTyPr 1091
 QY 2201 AGGGAGGCCA-----GAAGGCTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2236
 Db 1091 oGlyAsnProGlyIleLysGlySerValGlyAspProGlyLeuProGlyLeuProGlyTh 1111
 QY 2237 GGACTCTGCAAGTGGATTACCCAGGCTCAGGCTTAACAGCTAGCTCCTAGTTGAGAC 2296
 Db 1111 rProGlyAlaLysGlyGlnProGlyLeu-----1120
 QY 2297 ACCTAGAGAGGGTTTTTGGGAGCTGAATAAATACTCAGTCACTGCTGTTTCCATCTCT 2356
 Db 1121 -----ProGlyPheProGly-----ThrProGlyProProGlyProGly 1133
 QY 2357 CCCCTTAACCTGCACTTCTGTTAATGTAGCTTGTGCTGCTGCTGCTGCTGCTGCTG 2416
 Db 1133 s-----GlyIleSer-Gly-----1137
 QY 2417 CTCCTCCATGGATTGAACATATGACTTATTGTAGGGAAGAGTCTGCTGAGGGGCAACA 2476
 Db 1138 --ProProGlyAsn-----ProGlyLeuP 1145
 QY 2477 CACAAGAACAGGCTCCCTCAGCCACAGCACTGCTCTTTTGTGATCCACCCCTCTT 2536
 Db 1145 roGlyGluProGlyProValGlyGlyGlyGlyHisProGlyGlnProGlyPro-----1163
 QY 2537 ACCTTTTATCAGGATGTGGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2596
 Db 1163 -----1163
 QY 2597 TAAATATTTAACTATTATTTTAAAGTAAGGAATCCATGCTTGTGCTGCTGCTGCTG 2656
 Db 1163 -----1163

QY 312 AGCGGCTCACCCAGCCTCTGGACCATAGTGGG----- 278
 Db 684 oProGlySerProGlyGluGlnGlyPro-SerGlyAlaSerGlyProAlaGlyProArgG 704
 QY 277 -----CCAGCGGGTAGGCTCAGGGGCGCGTTCAGGCACCTCCAGAACTCCTCTGCTCTCG 223
 Db 704 lyProGlySerAlaGlySerProGlyLysAspGly-----LeuA 718
 QY 222 GCTCTGCTCCAGAGCTGCGGCCTCTCTCTGCTGCTGCCGCCAACTGCCTAGGAATCAGC 163
 Db 718 snGlyLeuProGlyProIleGlyProProGly-----ProArgGlyArgT 733
 QY 162 CAGCGCGCCATTCTGCCGCCCTTTGTCGCCGCTCCAGCTCTCTCAGCC 113
 Db 733 hrGly-----AspAlaGlyProAlaGlyProPro 742
 RESULT 22
 ID CA21_CANFA STANDARD; PRT; 1366 AA.
 AC 046392;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 2(I) chain precursor.
 DE COL1A2.
 GN Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RA Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC (FIBRILLAR FORMING COLLAGEN).
 CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 CC HYDROXYAPATITE.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF035120; AAC64485.1; -;
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib.collagen_C.
 DR Pfam; PF01391; Collagen; 17.
 DR Pfam; PF01410; COLFI; 1.
 DR ProDom; PD000007; Collagen; 2.
 DR ProDom; PD002078; Fib.collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 FT SIGNAL 1 24 POTENTIAL.
 FT PROPEP 25 79 AMINO-TERMINAL PROPEPTIDE (BY
 FT SIMILARITY).
 FT CHAIN 80 1102 COLLAGEN ALPHA 2(I) CHAIN.
 FT PROPEP 1103 1366 CARBOXYL-TERMINAL PROPEPTIDE (BY
 FT SIMILARITY).
 FT MOD_RES 80 80 PYRROLIDONE CARBOXYLIC ACID (BY
 FT SIMILARITY).
 FT MOD_RES 84 84 CONVERTED TO AN ALDEHYDE GROUP THAT IS
 FT INVOLVED IN CROSS-LINKING (BY
 FT SIMILARITY).

SQ SEQUENCE 1366 AA; 129400 MW; CD936969E080BCD4 CRC64;
 Alignment Scores:
 Pred. No.: 8.78e-08 Length: 1366
 Score: 321.00 Matches: 298
 Percent Similarity: 31.87% Conservative: 72
 Best Local Similarity: 25.67% Mismatches: 368
 Query Match: 5.16% Indels: 424
 DB: 1 Gaps: 64
 US-09-759-143-110 (1-3410) x CA21_CANFA (1-1366)
 QY 3074 GCTTCTGCTCCAGCTAGCTCCACACAGGCTTGTGGAGCTGGTGGGAAAGTTGGGGT 3015
 Db 227 AlaProGlyProAlaGlyAlaArg-----GlySerAspGlySer 239
 QY 3014 AGGGGAAGCTGGGGGTAGGGGAAATTTTGGGCAGTCG----- 2976
 Db 240 ValGlyProValGlyProAlaGlyProIleGlySerAlaGlyProGlyPheProGly 259
 QY 2975 -----TTCATCAGCCCGCAGCTCCTAGAGAGAGTAGAGGGGAGTGGAACT 2934
 Db 260 AlaProGlyProLysGlyGluIleGlyProVal-----GlyAsnProGly 274
 QY 2933 GGGGGGAACAGCTGGCGCCAGAGAAGAGGGGTGGTTAGGGAAGCCGTTGAGACCTCAA 2874
 Db 275 ProAlaGlyProAlaGlyProArgGlyGluValGlyLeuProGlyValSerGlyProVal 294
 QY 2873 GCCCCACCTCTACCTTCTTCAACACCCCTAACCTTGGGTAAACAGCATTTTGAATATCA 2814
 Db 295 GlyProProGly-----AsnProGlyAlaAsnGlyLeuThrGlyAla 308
 QY 2813 TTTGGGATGAGTAGAATTTCCAAAGTCTGGGTAGGCATTTTGGGGGCGCAGACCCAG 2754
 Db 309 LysGlyAlaAlaGlyLeuProGlyValAlaGly-----Ala-ProGlyLeuProGlyProAr 327
 QY 2753 GAGAAGAAGATTCTGGCAATGATCAGCCCAATGACCAGCTATCTCAGGGGACCTG----- 2699
 Db 327 g-----GlyIleProGlyProValGlyAl 335
 QY 2698 -----ATTGTTGGGATCCCCACCCCTACCCCAATATTAG 2664
 Db 335 aAlaGlyAlaThrGlyAlaArgGlyIleValGlyLeuProGlyProAlaGlySer----- 353
 QY 2663 ACACCAACACAGAAAAGCTAGCAATGATTCCTCTCTACTTTTAAATAAATAAGTTAA 2604
 Db 354 ----LysGlyGluSerGlyAsnLysGlyGluProGlySer----- 365
 QY 2603 ATATTTAAATGCTGTGTCTGTGTGTCGCAAGAGAGACCAAGAGCCGACATCTCTGAT 2544
 Db 366 -----AlaGlyAlaGlnGlyProGlyProSerGlyGly 377
 QY 2543 AAAAGGTAAGAGGGGGTGGATCAGCAAAAAGACAGTGTGT-----GG 2500
 Db 377 uGluGlyLysArgGlyProAsnGlyGlyAlaGlySerAlaGlyProSerGlyProProGly 397
 QY 2499 GCTCAGGGGA---CCTGGTCTTTGTGTGTTGCCCTCAGGACTCTTCCCTACAAATAAG 2443
 Db 397 yLeuArgGlySerProGlySerArgGlyLeuProGlyAlaAspGlyProAlaGlyValMe 417
 QY 2442 TCATATTTCAATCCCATGAGGAGGTTCATCTCTAGAACTCCCATCAGAGAGCTAC 2383
 Db 417 t-----GlyProProGlyProArgGlyAlaThrGlyProAlaGly 430
 QY 2382 ATTAACGAAGCTCAGGTTAAAGGGCTTAGAGATGGAAACACAGGT----- 2336
 Db 430 yValArgGlyProAsnGlyAspSerGly-ArgProGlyGluProGlyLeuMetGlyProA 450
 QY 2335 -----GACTGAGTTTATTCAGTCCCAAAACCCCTTCTCTAGGTG 2296
 Db 450 rgGlyPheProGlyAlaProGlyAsnValGlyProAlaGlyLysGluGlyProMetGlyL 470
 QY 2295 TGTCTCAACTAGGAGGCTAGCTGTTAACCCCTGACCCCTGGGTAAATCCACCTGCAGAGTCCC 2236

Db 470 euProGlyIleasp --- GlyArgProGlyProIleGlyP 482
QY 2235 CGCATTCAGTCAGGAGCCCTTCTGCGCTCCCTGTATAGTCCAGACTGAACCCCT 2176
Db 482 roAlaGlyAlaArgGlyGluProGlyAsnIleGlyPhe ---ProG 496
QY 2175 TGGAAAGGCTCCAGTCAGGACGCCCTAGAGACTGGGGAGAGAG --- 2132
Db 496 lyProLysGlyProThrGlyAspProGlyLysAsnGlyAspIleGlyHisAlaGlyLeuA 516
QY 2131 -----AGAGGAGCGCCAGCCAGCCAGCTGTCAGCTACGACCTCAGCAGCAGG 2080
Db 516 laGlyAlaArgGlyAlaProGlyProAspGly ---AsnAsnGlyAlaGlnG 532
QY 2079 GTGGCAGCAGAGACCATATTCTTGGCAGCAACAGAAACTGGCGCCAGCCGCGGAGC 2020
Db 532 ly-----ProProGlyP 536
QY 2019 CCATATGGGCT---AACAGAGCGGGAGCTGGGACCCAGTGGAGCGCCCTCCA --- 1967
Db 536 roGlnGlyValGlnGlyLysGlyGlnGlnGlyPro-----AlaGlyProProGlyP 554
QY 1966 -----CCCCAATGTCTGGAAGTTTCTACGCTGAGTATTGGCCAACTCGCTC 1918
Db 554 heGlnGlyLeuProGlyProAlaGlyThr ---Lys-----AlaGlyGluValGly 568
QY 1917 TTGTCAATACTACTGTGTAGCAAAAGTAAATGTCGACAGCCAGCCGCTGGCGAGC 1858
Db 569 -----Lys-----ProGlyGluArgGlyLe 575
QY 1857 ACCATATAGGAGTACAGACTGCTGAGCTGGACATGGAGCCCAATAACAGGATGGG 1798
Db 575 uProGly-----GluPheGlyLeuProGlyProAlaGlyProArgGlyGluArgG 592
QY 1797 GCCACCTGGGACAGCAGGACTATCCA -----GGATGGCGAGTCCAGG 1750
Db 592 yProProGlyGluSerGlyAlaAlaGlyProSerGlyProIleGlySerArgGlyProse 612
QY 1749 CAGATGCCCGCCCGGAAC -----CACCTGGCTCGGTGGCTCAGCCACC 1702
Db 612 rGlyProProGlyPro-AspGlyAsnLysGlyGluProGlyVal-LeuGlyAlaProGly 631
QY 1701 ACCACAGTACGGAGATCACAGGCAGAGCCCGCCGAGCCGCGGTGAGTGGAGC 1642
Db 632 Thr-----AlaGlyAlaSerGlyPro 638
QY 1641 AGGCCACTGCTCCAGCAGCCAGCTGTTCATTAGGGAAGGAGCTCCAGGCTTAGGCGCT 1582
Db 639 GlyGlyLeuProGlyGlu-----ArgGlyAlaAlaGlyIle---Pro 651
QY 1581 GGCAGGAGCTGGTCATCAGGCTGCTCTACTGCTAGCACCTCCAGTGTCCCTCGGTAT 1522
Db 652 Gly-GlyLysGlyGlyLysGly -----GluThrGlyLeuArgG 664
QY 1521 TTGGGACGAGAACACTTCTCCCGTGTGTAGAGGAGCCAGTGTGTAGGCGAGATC 1462
Db 664 yGluIleGlyAsnProGlyArgAspGlyAlaArgGlyAlaPro-----GlyAl 680
QY 1461 TGCAGGCTCAGAAGGTGAACCCGGTGAGGGCGCTGAAGCTGTCCACCGCCAGCAGCTG 1402
Db 680 aMetGly-----AlaProGlyProAlaG 688
QY 1401 TGGGACAGGATGTGGCAGCGGAGCCAGCCAGGAAAGCTGGCCACTGGCCAAATAGACT 1342
Db 688 yAlaThrGly-----AspArgGlyGly----- 695
QY 1341 GCTCGAGTGCAGGAATCGTCACAGCCGCTCATCATCAGAGAGAGACCGAGATG 1282
Db 696 -----AlaGlyProAlaGlyPro----- 701
QY 1281 GCGACTGTCAGGAACAGCC-----CCAGGCTGCCCATCCGAAACGCTTTCATCATAGTCT 1228
Db 1281 -----ArgGly-----LeuProGlyLeuLys 1023

702 -----AlaGlyProAlaGlyProArgGlyThrProGlyGluArgGlyGluValGlyP 719
1227 CTCGGGCTCTCGG -----TGCCCGCTCAGCTCTGGCGCACGCCCTGTACTA --- 1182
Db 719 oAlaGlyProAsnGlyPheAlaGlyProAlaGlyAlaAlaGlyGlnProGlyAlaLysG 739
1181 -----GCCCTCTCGC -----CCAGAAATCCGTG 1159
Db 739 yGluArgGlyThrLysGlyProLysGlyGluAsnGlyProValGlyProThrGlyProI 759
1158 TAAACACGCTGAAGGTTCATGTCATCCAGCTCAGCTGCACAGCTCAGCCAGAGACCCG 1099
Db 759 eGlySerAla -----GlyProSerGlyProAsnGlyProProGlyProAlaG 775
1098 CGCA -----GGTGGCGGCATCGCGCACAGCT -----GGTGCAGCCCGGGAAGC 1051
Db 775 ySerArgGlyAspGlyProGlyAlaThrGlyPheProGlyAlaAlaGlyArgTh 795
1050 AGGCGCGCCAGGT -----TCCGGAAGCCAGGCGCCGCGCATGGACAGCAG 1003
Db 795 rGlyProProGlyProSerGlyIleThrGlyProProGlyProProGlyAlaAlaGly 815
1002 TGGGCGCACAAGG -----AGGGGCGCACAGCCCTTCTGCTGCTCGCTCGGTG 958
Db 815 sGluGlyLeuArgGlyProArgGlyAspGlnGlyPro-----ValGlyArgTh 831
957 GGGCCCGCAGCTCTCTCAGCCAGCAGCTGTGGCTGTCTACGAGCTGAGGAGATG 898
Db 831 rGlyGluThrGlyAlaSerGlyProGlyPheThr -----GlyGlyL 846
897 AGGTGAGCAGCCCAAGAGGCACT -----CCTCTGGTGGTCCCAAGT --- 855
Db 846 sGlyProSerGlyGluProGlyThrAlaGlyProProGlyThrProGlyProGlnGlyLe 866
855 ----- 855
Db 866 uLeuGlyAlaProGlyIleLeuGlyLeuProGlySerArgGlyGluArgGlyLeuProG 886
854 -----AG 853
Db 886 yValAlaGlySerValGlyGluProGlyProLeuGlyIleAlaGlyProProGlyAlaAr 906
852 GGGCGCAGGCGCACTGTGTCTCCAGTCAATGGCAGCAGGAGTAGCCAGCCAGCCCA 793
Db 906 gGlyPro-----ProGlyAlaValGlyAlaProGlyValAsnGlyAlaProG 922
792 AGACTGATCATGAAGCATAGACAGTAGGCTCGGCACAGTGTGTGGTCCGCGGAAG 733
Db 922 yGlu-----AlaGlyArgAspGlyAsnProGlyAsnAspGlyPro---ProGlyAr 938
732 AGGTGAGCAGCAGGCGCTCCAGTGGAGTGAAGCAGCAGCTGCCACAGAGTCCAGCAGC 673
Db 938 gAspGlyGlnAlaGlyHisLysGlyGluArgGlyTyrProGlyAsnIleGlyProValG 958
672 CCCAGCCCGCAGGATGAGTCCAGCTCCAGGCGCTCGGATCCGCGGCAGCAGCAGCCCT 613
Db 958 y-----AlaValGlyAlaProGlyProHisGlyPro----- 968
612 GCTACCGCAGCGCGCTTGGGATGAGAAGAGGCTCAGCAGATGCCCAAGACAGTGC 553
Db 969 -----ValGlyProThrGly---LysHisGlyAsnArgGlyGluPro-GlyProAlaG 985
552 CAGATGAAGCGCGCGCGCTCCAGTGGTCCAGCCAGTGGTCACTGGCTGAGCCTAGG 493
Db 985 lySerValGlyProValGlyAlaValGly-ProArgGlyProSerGlyProGlnGlyLe 1004
492 AGCGGACACAGACCGCC ---AGCAGTGGACCAATGCCAGCAGCCATGTCATGAAC 436
Db 1005 ArgGlyAspLysGlyGluProGlyGlnLysGlyPro----- 1016
435 TTCCTCTACCCCGCTTCCAGCAGCAGGCGGCACATAGTGTGCTGCGCGCAAA 376
Db 1017 -----ArgGly-----LeuProGlyLeuLys 1023

QY 375 CACACCTCCAGCCAAAGGTTGACGAGTTCACACAGAGAGCTGGCGTTTCGGT---GC 319
 Db 1024 GlyHisasn-Gly-----LeuGlnGlyLeuProGlyLeuAlaGlyClnHisGlyAspG 1041
 QY 318 CGCAGCAGCGGCTCACCCACAGCCTCTGACCATAGTGGCCAGCGCGG----- 269
 Db 1041 nGlyAlaProGlySerValGlyProAlaGlyPro-ArgGlyProAlaGlyProSerGlyP 1061
 QY 268 --TAGGCTCAGGGCGGCTTACAGGACTCCAGAACTG-----CITCGTC 226
 Db 1061 roAlaGlyLysAspGlyArgThrGlyClnProGlyThrValGlyProAlaGlyLeuArg 1081
 QY 225 TCGCTCTGCTCAGAGAGCTCGCGCTCTCTCTGCTGCTCGCCCAACTGCCTAGGAATC 166
 Db 1081 lySerGlnGlySerGlnGlyProAlaGlyProGly-----ProProGlyP 1097
 QY 165 AGCCAGCGCCCAATTTCTGCGAGCCCTTTGGTCCCGT----- 128
 Db 1097 roProGlyProPro-----GlyProSerGlyGlyGlyTyrAspPheGlyTyrGluGlyA 1115
 QY 127 -----CCAGCTCTCAGCCCATGCTC 107
 Db 1115 spPheTyrArgAlaAspGlnProArgSerProProSerLeu 1128

RESULT 23

CA17_HUMAN STANDARD; PRT: 2944 AA.
 AC Q02388; Q14054; Q16507;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(VII) chain precursor (Long-chain collagen) (LC collagen).
 GN COL7A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94327588; PubMed-8051117;
 RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;
 RT "Cloning of human type VII collagen. Complete primary sequence of the alpha 1(VII) chain and identification of intragenic polymorphisms.";
 RL J. Biol. Chem. 269:20256-20262(1994).
 RN [2]
 RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE-93338437; PubMed-1307247;
 RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G., Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E., Uitto J.;
 RT "The large non-collagenous domain (NC-1) of type VII collagen is amino-terminal and chimeric. Homology to cartilage matrix protein, the type III domains of fibronectin and the A domains of von Willebrand factor.";
 RL Hum. Mol. Genet. 4:475-481(1995).
 RN [3]
 RP SEQUENCE OF 815-1439 FROM N.A.
 RX MEDLINE-91334380; PubMed-1871109;
 RA Parente M.G., Chung L.C., Ryyanen J., Woodley D.T., Wynn K.W., Bauer E.A., Mattei M.-G., Chu M.-L., Uitto J.;
 RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).
 RN [4]
 RP SEQUENCE OF 369-1255 FROM N.A.
 RX MEDLINE-93107742; PubMed-1469284;
 RA Gammon W.R., Abernethy M.L., Padilla K.M., Prisanan P.S., Cook M.E., Wright J., Brigaman R.A., Hunt S.W. III;
 RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion proteins involved in tissue-specific organization of extracellular matrix.";

RL J. Invest. Dermatol. 99:691-696(1992).
 RN [5]
 RP SEQUENCE OF 340-675 FROM N.A.
 RX TISSUE-Keratinocytes;
 RC MEDLINE-92231902; PubMed-1567409;
 RA Tanaka T., Takahashi K., Furukawa F., Inamura S.;
 RT "Molecular cloning and characterization of type VII collagen cDNA.";
 RL Biochem. Biophys. Res. Commun. 183:958-963(1992).
 RN [6]
 RP SEQUENCE OF 2395-2944 FROM N.A.
 RX MEDLINE-93271985; PubMed-8499916;
 RA Greenspan D.S.;
 RT "The carboxyl-terminal half of type VII collagen, including the non-collagenous NC-2 domain and intron/exon organization of the corresponding region of the COL7A1 gene.";
 RL Hum. Mol. Genet. 2:273-278(1993).
 RN [7]
 RP SEQUENCE OF 1-87 FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE-94375010; PubMed-8088784;
 RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W., Uitto J., Greenspan D.S.;
 RT "Structural organization of the human type VII collagen gene (COL7A1), composed of more exons than any previously characterized gene.";
 RL Genomics 21:169-179(1994).
 RN [8]
 RP REVIEW ON DEB VARIANTS.
 RX MEDLINE-98041696; PubMed-9375848;
 RA Jaervikallio A., Pulkkinen L., Uitto J.;
 RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in the type VII collagen gene (COL7A1).";
 RL Hum. Mutat. 10:338-347(1997).
 RN [9]
 RP VARIANT RDEB LYS-2798.
 RX MEDLINE-93291877; PubMed-8513326;
 RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y., Lin A.N., Dietz H.C., Hovnanian A., Uitto J.;
 RT "A missense mutation in type VII collagen in two affected siblings with recessive dystrophic epidermolysis bullosa.";
 RL Nat. Genet. 4:62-66(1993).
 RN [10]
 RP VARIANT DDEB SER-2040.
 RX MEDLINE-94224777; PubMed-8170945;
 RA Christiano A.M., Ryyanen M., Uitto J.;
 RT "Dominant dystrophic epidermolysis bullosa: identification of a Gly-->Ser substitution in the triple-helical domain of type VII collagen.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).
 RN [11]
 RP VARIANT PEB-DDEB CYS-2623.
 RX MEDLINE-96081220; PubMed-8541842;
 RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Uitto J.;
 RT "Pretibial epidermolysis bullosa: genetic linkage to COL7A1 and identification of a glycine-to-cysteine substitution in the triple-helical domain of type VII collagen.";
 RL Hum. Mol. Genet. 4:1579-1583(1995).
 RN [12]
 RP VARIANT DDEB ARG-2043.
 RX MEDLINE-95164985; PubMed-7861014;
 RA Christiano A.M., Morricone A., Paradisi M., Angelo C., Mazzanti C., Cavallieri R., Uitto J.;
 RT "A glycine-to-arginine substitution in the triple-helical domain of type VII collagen in a family with dominant dystrophic epidermolysis bullosa.";
 RL J. Invest. Dermatol. 104:438-440(1995).
 RN [13]
 RP VARIANTS RDEB AND DDEB.
 RX MEDLINE-96220218; PubMed-8644729;
 RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;
 RT "Glycine substitutions in the triple-helical region of type VII collagen result in a spectrum of dystrophic epidermolysis bullosa phenotypes and patterns of inheritance.";
 RL Am. J. Hum. Genet. 58:671-681(1996).

1349 ProGlyAlaProGlyGlnValIleGlyGlyGluGlyProGlyLeuProGlyArgLysGly 1368
1745 TGCCCGGCGCCGAA---CCACCTCGCTCGGTGGCTCACCCACCACACACGTACGG 1689
1369 AspProGlyProSerGlyProGlyProArg-Gly- 1383
1688 AGACATCACAGGAGGCGCC---GCAGCGCGG 1656
1383 YAspProGlyProArgGlyProGlyLeuProGlyThrAlaMetLysGlyAspLysGly 1403
1655 GTGAGGTGGAGCAGCCACTCCCTCCAGCACCCAGCTGTCCATTAGGGAAGGAGCTC 1596
1403 YAspArg-GlyGluArgGly---ProProGlyPro- 1416
1595 CAGGCTTAGGCTGGCAGGAACTGTCTCATCAGGCTGTCTCACTGTAGCACCTCCAG 1536
1416 lYgIleAlaProGly- 1421
1535 TGTCCTCCCTCGTATTTGGGAGGAAACACTGCTTCTCCGGTGTGTAGAGGAGGCGCAGTG 1476
1422 --GluPro-GlyLeu---ProGlyLeuProGlySerProGlyProGlyGly---ProVal 1438
1475 TGTAGGCGAGTCTCGAGGCTGAGAGGTGAACCCGGTGAGGCGGCTGAAGCTGTCA 1416
1439 GlyProProGlyLysLysGlyGlyLysGlyAspSerGluAspGlyAlaProGlyLeuPro 1458
1415 CCAGCGCCACACTGTGGGACAGGCTGTGCACCGCGCAGCCAGGAAAGCTG- 1362
1459 GlyGlnProGlySerProGlyGluGlnGlyProArgGlyProProGlyAlaIleGlyPro 1478
1361 -----CCACACTGGCCAAATAGACTCTCGAGTGGCGAATCTCGTCCACGACCGGT 1311
1479 LysGlyAspArgGlyPheProGlyProLeuGlyGluAlaGlyGlyLysGlyGluArgGly 1498
1310 CCATGACAGAGAGAACAGGAGATGCGGCACCTCCAGGAACAGCCCGGCTGCCCA 1251
1499 Pro-----ProGly-----ProAlaGlySerArg---GlyLeuPro 1509
1250 TCCGAACGCTTCATAGTGTCTCGCGGCTCGTCCCGGCTCAGCTCTGGGACGC 1191
1510 GlyValAlaGlyArgProGlyAlaLysGlyProGlyGlyProGlyProThrGlyArg 1529
1190 CCTGGTACAGCCCTCCCGCAGCAATTCGTGTAAACAGCGTGAAGTTCATAGTGCCA 1131
1530 GlnGlyGlyLysGlyGlyProGlyArgProGlyAspProAla-----ValVal 1545
1130 TCCAGCTGCACAGTCCAGCAGGAGAGCGCGGCGAGGTGCGGGCATCGGC- 1077
1546 GlyProAlaValAlaGlyProLysGlyLysGlyAspValGlyProAlaGlyProArg 1565
1076 AGCAGCTGTGTGCGCGGAGAGCAGGCGCGCCAGTTCGCGAAGCCAGCGGCGCC 1017
1566 GlyAlaThrGlyValGlnGlyGluArgGlyProProGlyLeuValLeuProGlyAspPro 1585
1016 GGCATGACAGCTGGGCGGCACAGAGGCGGCGGCGGCTCTCTGCTGGCTCGGTGG 957
1586 GlyProLysGlyAspProGlyAsp---ArgGlyProIleGly---LeuThrGlyArgAla 1603
956 GCGCCAGCGCTGCCT-----CCT 939
1604 GlyPro-----ProGlyAspSerGlyProProGlyGlyLysGlyAspProGlyArgPro 1621
938 CAGCCACAGCAGTGTGCTGCTACGAGGTGAGAGATGAGGCTGAGGAGGCGCAAGA 879
1622 GlyProProGlyProValGlyProArg-----GlyArgAspGlyGluValGlyGluLys 1639
878 GGCAC-----CCTCTGGGTGCCAGGT----- 855
1640 GlyAspGlyGlyProProGlyAspProGlyLeuProGlyLysAlaGlyGluArgGlyLeu 1659
854 -----AGGGGCGCA----- 828
1660 ArgGlyAlaProGlyValArgGlyProValGlyGlyLysGlyAspGlnGlyAspProGly 1679

827 CAATGGCAGGAGGAGGTAGCCAGGAGAGCCCAAGACTGATCATGAAGCATAGACAG 768
1680 GluAspGlyArgAsnGlySerProGlySerGlyPro-----LysGlyAspArg 1696
767 AGTAGGCTGCGACAGTGGTCCGGGTCCCGAAGAGAGTCAAGAGAGGCGCTCCAGTG 708
1697 GlyGluProGlyProProGlyPro---ProGlyArgLeuValAspThrGlyProGlyAla 1715
707 ---GAGTGAAGCACACCTGGCCACAGAGTCCAGCAGGCCCCACGC----- 666
1716 ArgGlyLysGlyGluProGlyAspArgGlyGlnGlyProArgGlyProLysGlyAsp 1735
665 CCAGGATGAGCAGTCCAGCTCCAGGCGCTGGATCCGGCAGCAGCAGCCCTCTAGCC 606
1736 ProGly-----LeuProGlyAla-----ProGly----- 1743
605 AGCGCGCTTGGGTGATGAAGAGAGGCTCAGCAGGATGC-----CCAAGG 561
1744 -----GluArgGlyIleGlyPheArgGlyProProGlyProGln 1757
560 ACAGTCCCGCAGATGAAGGCGCGGCATAGCTCCACGCGCTGCTACTGGCTG 501
1758 GlyAspProGlyValArgGlyProAlaGlyGlyLys----- 1769
500 AGCTTAGGAGCGGACACAGAGCCAGCAGCTGGACCAATGCCAGCAGCAGTGTCA 441
1770 -----GlyAspArgGlyProProGlyLeuAspGly----- 1779
440 TGAACCTCTCTCTACCCCTCCAGCAGCAGAGCGGCACATAGTGTGCTGCTGG 381
1780 ---ArgSerGlyLeuAspGlyLysProGlyAlaAlaGly----- 1791
380 CCAACACACACTCCAGGCAAGG---TTAGCAGGTGACAGCAGAGCTGGCTTTC 324
1792 ProSerGlyProAsnGlyAlaGlyLysAlaGlyAspProGlyArgAspGlyLeuPro 1811
323 GGT---GCCGACAGGCGGCTCAGCAGCAGCTCTGACCATAGTGG----- 278
1812 GlyLeuArgGlyGluGlnGlyLeuProGlyProSerGlyPro-ProGlyLeuProGlyLys 1831
277 -----CCAGGCGGTGAGGCTCAGGCGCGCTTCAGGCGCTCCA--- 239
1831 sProGlyGluAspGlyLysProGlyLeuAsnGlyLysAsnGlyGluProGlyAspProGly 1851
238 -GAAGTGTCTCTCGCTCTGCTTCCAGAGCTGGGCTCTCTCTCTGCTGCGGCA 180
1851 yGluAspGlyArgLysGlyGlyLysGlyAspSerGlyAlaSerGlyArgGlyArgAs 1871
179 ACTGCTAGGAATCAGCCAGCGCC---CATTTCTGCCAGCCCTTTGGTGGGCTCCAGC 123
1871 pGlyProLysGlyGluArgGlyAlaProGlyIleLeuGlyProGlnGly-ProProGlyLys 1891
122 TTCTCAGCCCATGCTCAACACCTGCT---GCTGTGGGCGACCTCAGTGGGAGACGCTCTC 66
1891 euProGlyProValGlyProProGlyGlnGlyPheProGlyValProGlyLysThrGlyP 1911
65 ATCACTCAGATCTGCTGCT-----CGAG 45
1911 rolyGlyAspArgGlyGluThrGlySerLysGlyGlnGlyLeuProGlyGluArgG 1931
44 GCGCGCGGTGTCCACCGGAGCC 22
1931 lYleuArgGlyGluProGlySer 1938
RESULT 24
CA24_ASCSU
-ID CA24_ASCSU STANDARD; PRT; 1763 AA.
AC P27393;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.

OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 OC Ascariidae; Ascaris.
 OX NCBI_TaxID=6253;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91340768; PubMed=1714907;
 RA Pettitt J., Kingston I.B.;
 RT "The complete primary structure of a nematode alpha 2(IV) collagen
 and the partial structural organization of its gene.";
 RL J. Biol. Chem. 266:16149-16156(1991).
 CC -!- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
 CC -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
 CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1
 CC DOMAINS.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M67507; AAA18014.1; .
 CC PIR; S16366; S16366.
 CC DR InterPro; IPR000087; Collagen.
 CC DR InterPro; IPR001442; ProcollagenC4.
 CC DR Pfam; PF01391; Collagen; 25.
 CC DR Pfam; PF01413; C4; 2.
 CC DR ProDom; PD000007; Collagen; 3.
 CC DR ProDom; PD003923; ProcollagenC4; 2.
 CC DR SMART; SM00111; C4; 2.
 CC KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
 CC Alternative splicing; Glycoprotein; Signal.
 CC SIGNAL 1 26
 CC CHAIN 27 1763
 CC COLLAGEN ALPHA 2(IV) CHAIN.
 CC 7S DOMAIN.
 CC TRIPLE-HELICAL REGION.
 CC NONHELIICAL REGION (NC1).
 CC OR 1634 (BY SIMILARITY).
 CC OR 1637 (BY SIMILARITY).
 CC BY SIMILARITY.
 CC OR 1749 (BY SIMILARITY).
 CC OR 1752 (BY SIMILARITY).
 CC BY SIMILARITY.
 CC N-LINKED (GLCNAC...) (POTENTIAL).
 CC O-LINKED (GLYCOSAMINOGLYCAN) (IN FORM II).
 CC (POTENTIAL).
 CC GEOPRGPQGPVPSVGAAGTIIIGEPAPGMKEK ->
 CC IDIGPAGPPGPPGPPFTGSGSIVGPRHSGDSKGVK (IN
 CC ISOFORM II).
 CC VARSPLC 230 266
 CC SEQUENCE 1763 AA; 168526 MW; 304F528BC06AAE0D CRC64;
 Alignment Scores:
 Pred. No.: 2,04e-07 Length: 1763
 Score: 312.50 Matches: 315
 Percent Similarity: 31.65% Conservative: 60
 Best Local Similarity: 26.58% Mismatches: 430

Query Match: 4.87% Indels: 381
 DB: 1 Gaps: 68
 US-09-759-143-110 (1-3410) x CA24_ASCSU (1-1763)
 QY 2 GAACACAGCCTGCAGCGCTGCTCGGGTGACAGCGCGCGCTCGGCAGGATCTGAG 61
 DB 404 GlyProGlyLeuProGlyLeu---ProGlyLeuGluGlyLeuProGlyProGlyGlyGlu 422
 QY 62 TGATGAGACCTGTCCCGCCTGAGGTGCCCGACAGCAGCAGGTGTTGAG-----CATGGG 115
 DB 423 LysGlyAspSerGly-----IleProGlyAlaProGlyValGlnGlyProGly 439
 QY 116 CTGAGAAGCTGACCGCGCACCAAGGCTGGCAGAAATGGCCGCT-----GG 163
 DB 440 LeuAlaGlyProProGlyAlaLysGly-----GluProGlyProArgGlyValAspGly 457
 QY 164 CTGATTCCTAGCAGTTGGCGCAGCAAGAGGAGGAGCGCAGCTCTTGAGCAGAGCC 223
 DB 457 yGlnSerIleProGlyLeuProGlyLysAspGlyA-gProGlyLeuAspGlyLeuProGly 477
 QY 224 GAGACGAAGCAGTTCTGGAGTGCCT-----GAACGGCCCGCTGAG-----263
 DB 477 yA-gLysGlyGluMetGlyLeuProGlyValA-gGlyProProGlyAspSerLeuAsnGly 497
 QY 264 ----CCCTACCGCCCTGGCCCGCCTATGCTCCAGAGGCTGTGGGTGAGCCGCTGTCGGG 319
 DB 497 yLeuProGlyProProGlyPro-ArgGlyProGln-----G 509
 QY 320 CACCGGAA-----AGCCGACGCTCTTGCTGCTCAACCTGCTAACC 358
 DB 509 lyProLysGlyTyArgAspGlyArgAspGlyAlaProGlyLeuProGlyIlePro-----526
 QY 359 TTTGGCTGGAGGTGTTGGCGCAGGACATCCTATGCTGCCGCTCTGCTGCTGGAA 418
 DB 527 -----GlyProLysGlyAspArgGlyGlyThrCysAlaPheCysAla-----H 541
 QY 419 GTGGGGTAGAGCAGAGTTTCATGACCATGGTCTGGCATTTGGTCCAGCTGCTGGGCT- 477
 DB 541 isGlyAlaLysGlyGluLysGlyAsp-----AlaGlyTyArgGlyLeuProGlyProG 559
 QY 478 -----GGTC 481
 DB 559 lnglyGluArgGlyLeuProGlyIleProGlyAlaThrGlyAlaProGlyAspAspGlyL 579
 QY 482 TGTGTCGCTCTAGGCTAGCAGCAGTACCATGCGGTGGAGCTATGGCGCCGCCGG 541
 DB 579 euProGlyAlaProGlyArgProGlyPro-----GlyProProGlyG 594
 QY 542 CCCTTCATCTGGGCACTGCTTGGGCATCCTGCTGAGCCTCTTTCTCATCCCAAGGGCC 601
 DB 594 lnAspGlyLeuProGlyLeuProGlyGlnLysGlyGluProThr-----G 609
 QY 602 GGTGCTAGCAGGCTGTGTCGCGGATCCCGAGGCCCTGGAGCTGGCAGCTGCTCATC 661
 DB 609 lnLeuThrLeuArgProGlyProProGlyTyArgProGlyGlnLysGlyGluThrGlyPheP 629
 QY 662 CTGGG---CGTGGGCTGTGACTTCTGTGGCCAGGTGTGTTCACTCCACTGAGAGGCC 718
 DB 629 roGlyProArgGlyGlnGluGlyLeu-----ProGlyLysProGlyIleValGlyAlaP 647
 QY 719 CTGCTCTGACCTCTTCGGGA---CCCGGACCACTG---TCGCCAGGCTACTCTGTC 772
 DB 647 rogly-----LeuProGlyProProGlyGlyGluProGlyLeuThrGlyL 664
 QY 773 TATGCTTCATGATCAGTCTTGGGGCTGCTGCGCTACCTCTGCTGCTGCTGCTGCTGCTG 832
 DB 664 euProGlyLysProGlyLysAspGlyIleProGlyLeuProGlyLeuLysGlyGluProG 684
 QY 833 GACACAGTGCCTGCGCCCTTACCTGGGACCCAGGA-----GGAGTGCCTCTTTGTC 886
 DB 684 lyTyrglyGlnProGly---MetProGlyLeuProGlyMetLysGlyAspAlaGlyLeuP 703

Db 1266 lnProGlyLysSerIleThrGlyProLysGlyAsnAlaGly-LeuProGlyLeuProGly 1285
 QY 2824 AAATGCTGTACCAAGGTAGGCTGTGAAGNAGGTAGAGGT----- 2868
 Db 1286 LysAspGlyLeuProGlyLeuProGlyLeuLysGlyGlyLeuProGlyLysProGlyTyrAla 1305
 QY 2869 -----GGGGCT 2874
 Db 1306 GlyAlaAlaGlyIleLysGlyGluProGlyLeuProGlyIleProGlyAlaLysGlyGlu 1325
 QY 2875 TCAGGTCTCAAGGCTCCCTAACCCAC-----CCCTCTCTCTGTGGC 2916
 Db 1326 ProGlyLeuSerGlyIleProGlyLysArgGlyAsnAspGlyIleProGlyLysProGly 1345
 QY 2917 CCAGCCGTGGTCCGCCACATCCA-----CTCCCTCTACTCTC 2955
 Db 1346 ProGlyLeuProGlyLeuProGlyMetLysGlyGluSerGlyLeuProGlyProGln 1365
 QY 2956 TCTAGGACTGGCTGATGAAGGCACCTGCCCAAAATTTCCCTACCCCAACTTCCC 3012
 Db 1366 GlyProAlaGlyLeu---ProGlyLeuProGlyLeuLysGlyGluProGlyLeuPro 1383
 RESULT 25-
 CA25_HUMAN STANDARD; PRT; 1496 AA.
 AC P05997;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 2(V) chain precursor.
 GN COL5A2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-463 FROM N.A.
 RX MEDLINE=89123368; PubMed=2914927;
 RA Woodbury D., Benson-Chanda V., Ramirez F.;
 RT "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
 RT to the structural criteria of a fibrillar procollagen molecule.";
 RL J. Biol. Chem. 264:2735-2738(1989).
 RN [2]
 RP SEQUENCE OF 398-1496 FROM N.A.
 RX MEDLINE=87146331; PubMed=3029669;
 RA Weil D., Bernard M.P., Gargano S., Ramirez F.;
 RT "The pro alpha 2(V) collagen gene is evolutionarily related to the
 RT major fibrillar-forming collagens";
 RL Nucleic Acids Res. 15:181-198(1987).
 RN [3]
 RP SEQUENCE OF 1227-1496 FROM N.A.
 RX MEDLINE=85289337; PubMed=2411731;
 RA Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
 RT "Complete primary structure of the human alpha 2 type V procollagen
 RT COOH-terminal propeptide";
 RL J. Biol. Chem. 260:11216-11222(1985).
 RN [4]
 RP SEQUENCE OF 1449-1496 FROM N.A.
 RX MEDLINE=89138450; PubMed=3224983;
 RA Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
 RA Ramirez F.;
 RT "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
 RT located on the long arm of human chromosome 2.";
 RL Genomics 3:275-277(1988).
 RN [5]
 RP SEQUENCE OF 208-227.
 RX MEDLINE=92239022; PubMed=1571108;
 RA Mann K.;
 RT "Isolation of the alpha 3-chain of human type V collagen and
 RT characterization by partial sequencing";
 RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).
 RN [6]

RP SEQUENCE OF 288-297 AND 606-617.
 RC TISSUE=Bone;
 RX MEDLINE=94237164; PubMed=8181482;
 RA Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champliand M.F.,
 RA Bouthillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
 RT "Diversity in the processing events at the N-terminus of type-V
 RT collagen";
 RL Eur. J. Biochem. 221:987-995(1994).
 RN [7]
 RP DISEASE.
 RX PubMed=9425231;
 RA Michalickova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;
 RT "Mutations of the alpha2(V) chain of type V collagen impair matrix
 RT assembly and produce Ehlers-Danlos syndrome type I.";
 RL Hum. Mol. Genet. 7:249-255(1998).
 RN [8]
 RP VARIANT EDS-II ARG-960.
 RX PubMed=9783710;
 RA Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,
 RA Burrows N.P.;
 RT "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type
 RT II.";
 RL J. Med. Genet. 35:846-848(1998).
 CC -!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE
 CC COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS
 CC TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.
 CC -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN
 CC MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND
 CC ONE ALPHA 3(V) CHAINS IN PLACENTA.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
 CC gravis or type I (EDS-I), and Ehlers-Danlos syndrome mitis or type
 CC II (EDS-II). Ehlers-Danlos syndrome is a genetically and
 CC phenotypically heterogeneous connective-tissue disorder
 CC characterized by loose-jointedness and fragile, velvety,
 CC stretchable, bruisable skin that heals with peculiar 'cigarette-
 CC paper' scars. EDS-I and EDS-II are autosomal dominant traits.
 CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04478; AAA51859.1; -
 DR EMBL: X04758; CAA28454.1; -
 DR EMBL: M11718; AAA52058.1; -
 DR PIR: A25374; A25374.
 DR PIR: A25874; A25874.
 DR PIR: A30017; A30017.
 DR PIR: A31427; A31427.
 DR Genew: HGNC:2210; COL5A2.
 DR MIM: 120190; -
 DR MIM: 130000; -
 DR MIM: 130010; -
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR000885; Fib.collagen_C.
 DR InterPro: IPR001007; VWFC.
 DR Pfam: PF00093; vwc; 1.
 DR Pfam: PF01391; Collagen; 18.
 DR Pfam: PF01410; COLFI; 1.
 DR ProDom: PD000007; Collagen; 5.
 DR ProDom: PD002078; Fib.collagen_C; 1.
 DR SMART: SM00038; COLFI; 1.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS01208; VWFC; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;

KW Disease mutation. 26
 FT SIGNAL 1 1226
 FT CHAIN 27 1226
 FT PROPEP 1227 1496
 FT DOMAIN 39 97
 FT MOD_RES 290 290
 FT MOD_RES 293 293
 FT MOD_RES 296 296
 FT MOD_RES 608 608
 FT MOD_RES 614 614
 FT MOD_RES 960 960
 FT VARIANT 960 960
 FT CONFLICT 292 292
 FT CONFLICT 1418 1418
 FT CONFLICT 1438 1438
 FT CONFLICT 1460 1460
 FT CONFLICT 1496 1496
 FT SEQUENCE 1496 AA; 144720 MW; 82827C17A8644F5A CRC64;

Alignment Scores:
 Pred. No.: 2.18e-07 Length: 1496
 Score: 312.00 Matches: 226
 Percent Similarity: 30.80% Conservative: 37
 Best Local Similarity: 26.46% Mismatches: 255
 Query Match: 4.86% Indels: 336
 DB: 1 Gaps: 51

US-09-759-143-110 (1-3410) x CA25_HUMAN (1-1496)

QY 26 CCGGGTGC-----AGCGCGCGCCTCGG-----CCAGGA 55
 DB 242 ProGlyAspProGlyProMetGlyProGlyArgGlySerArgGlyProGlyProGly 261
 QY 56 TCTGAGTGTGAGAGCGTGTCCACACGAGG-----85
 DB 262 LysProGlyGluAspGlyGluProGlyArgGlyAsnProGlyGluValGlyPheAla 281
 QY 86 -----TGCCCCACAGCAGAGGTGTGAGCATGGGTGAGAACG 124
 DB 282 GlySerProGlyAlaArgGlyPheProGlyAlaProGlyLeuProGlyLeuGly 300
 QY 125 TGGACCGCCACCAAGGGCTG-----145
 DB 301 HisArgGlyHisGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 320
 QY 146 -----GCAGAAATGGCGCTCG-----CTGATTCCTAGGCAGTTG 181
 DB 321 GlyGluAlaGlyProThrGlyProMetGlyAlaMetGlyProLeuGlyProArgGlyMet 340
 QY 182 CGGCACCAAGGAGGAGCGCGCTTCTGGAGCAGACCGCAGAGCAGGAGCAGTCTCGG 241
 DB 341 ProGlyGluArg-GlyArgLeuGlyProGlnGlyAlaProGlyGlnArgGlyAlaHisG 360
 QY 242 AGTCCCTGAACGGCC-----257
 DB 360 yMetProGlyLysProGlyProMetGlyProLeuGlyIleProGlySerSerGlyPhePr 380
 QY 257 -----257
 DB 380 oGlyAsnProGlyMetLysGlyGluAlaGlyProThrGlyAlaArgGlyProGlyGlyPr 400
 QY 258 -----CCTGAGCCCTACCGCCTGCCCACTATGT-----CCAG 292
 DB 400 oGlnGlyGlnArgGlyGluThrGlyProGlyProGlyProVal-GlySerProGlyLeuProG 420
 QY 293 AGGCTGTGGGTGAGCGCCCTGCTGCGCACCGGAA-----AGCCCAAGCTCTTG 340
 DB 420 lyAlaIleGlyThrAspGlyThrProGlyProLysGlyProThrGlySerProGlyThrS 440
 QY 341 CTGCTCAACCTGTAACTTTGGCGCTGGAGGTGTGTTGGCCGACGATCATCCTATGTG 400
 DB 440 erGlyProGlyLysAlaGlyProProGlySerProGlyProGln-----455

QY 401 CCGCCTCTGCTGCTGGAAAGTGGGGTAGAGAGAGAAAGTTTCATGACCATGGTCTGGGCATT 460
 DB 456 --GlySerThrGlyProGlnGlyAsnSerGlyLeuProGlyLys-----ProGlyPheL 473
 QY 461 GGTCCAGTGTGGGCTGTGTGTCCTCCCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTC 519
 DB 473 ysGlyGluAlaGlyProLysGlyGluProGlyProHisGlyIleGlnGlyProIleGlyP 493
 QY 520 -----TGGACCTATGGCCGCGC-----537
 DB 493 roProGlyGluGluGlyLysArgGlyProArgGlyAspProGlyThrLeuGlyProProG 513
 QY 538 ---CGGCGCCCTCATCTCTGGGCACTGTC-----CTTGGCATCCF---GCTGAGCCTCTTT 586
 DB 513 lyProValGlyGluArgGlyAlaProGlyAsnArgGlyPheProGlySerAspGlyLeuP 533
 QY 587 CTCATCCCAAGGCGCGCTGTGCTGCGCGGATGCTGCGCGGATCCAGCCCTCGGAG 646
 DB 533 roGlyProLysGlyAlaGlnGlyGluArgGlyProVal---GlySerSerGlyProLysG 552
 QY 647 CTGGCACTGCTCATCTCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
 DB 552 lySerGlnGlyAspProGlyArgProGlyGluProGlyLeu-----ProGlyAlaArgG 570
 QY 704 ACTCCACTGGAGGCCCTGCT-----CTCTGACCTCTTCCGGGAC 742
 DB 570 lyLeuThrGlyAsnProGlyValGlnGlyProGlyGlyLysLeuGlyProLeuGlyAlaP 590
 QY 743 CCGGA-----CCACTGTGCGCAGGCTACTCTGTCTATGCTTTCATGATCAGT 790
 DB 590 roGlyGluAspGlyArgProGlyProGly-----S 601
 QY 791 CTGTTGGGCTGCTGGGTACTCTCTGCTGCCATTGACTGCTGGAGACACAGTGGCCTGGCC 850
 DB 601 erIleIleIleLysGlyGlnProGlyThr-----MetGly-----LeuProGlyP 616
 QY 851 CCCTA-----CCTGGCACCCAGGAGGAGTCCCTTTTGGCCTGCTCACC 895
 DB 616 roLysGlySerAsnGlyAspProGlyLysProGlyGluAla-----GlyAsnP 632
 QY 896 CTCATCTCTCCACCTGCTGCTAGCAGCCACACTGCTGCTGGTGGAGGAGGAGCCTGGCC 955
 DB 632 roGlyValProGlyGlnArgGly-----AlaProGlyLysAspGlyLysValGlyP 649
 QY 956 CCCAC-----CGAGCCAGCAGAGGCTCTCGGCCCTCC 991
 DB 649 roTyrglyProProGlyProGlyLeuArgGlyGluArgGlyGluGlnGlyPro-----667
 QY 992 TTGTGCGCCCACTGTGTCCATGCGCGCGCTTGGCTTTCCGGAACCTCGCGCCCTG 1051
 DB 668 -----ProGlyProThrGlyPhe-----GlnGlyHisProG 678
 QY 1052 CTTCCCGGCTGCACACAGCTGTGTGCGCGCATGCCCGCACCCCTGCGCGGCTCTTCGTG 1111
 DB 678 lyProProGlyProGly-----684
 QY 1112 GCTGAGCTGTGACGCTGATGCACTCATGACCTTCACGCTGTTTACACGATTTCTGTG 1171
 DB 685 -----GluGly-----GlyLysProG 690
 QY 1172 GCGAGGGGCTGTACAGGGGCTGCCAGAGCTGAGCGGGCAGCGCGCGGAGACAC 1231
 DB 690 lyAspGlnGlyValProGlyGlyProGlyAla---ValGlyProLeuGlyPro-----706
 QY 1232 TATGATGAAGCGTTTCGATGGCAGCCTGG-----GCTGTTCTCTGATGCGGCCATC 1285
 DB 707 -----ArgGlyGluArgGlyAsnProGlyGluArgGlyGluProGlyIleThrGlyL 724
 QY 1286 TCCCT-----GGTCTTCTCTGCTGATGAGCGGCTGGTGCACGCTATCGGCCT 1336
 DB 724 euProGlyGluLysGlyMetAlaGlyHisGlyProAspGly-----738
 QY 1337 CGAGCAGTCTATTGCGCAGTGTGCGAGCTTTTCCCTGTGGTGTGCGGTCGCCACATGCTG 1396

```

Db      739  -----ProLysGly-----SerProG 744
QY      1397  TCCACAGTGGCGGTGGTGCACACCTTACCGCCCTCACCGGGTTCACCTTCACGCC 1456
Db      744  lyProSerGlyThrProGlyAspThr----- 752
QY      1457  CTGCAGATCCCTGCCTACACACTGGCTCCCTACACCGCGGAGACAGGCTTCTCTG 1516
Db      753  --GlyProProGlyLeuGln---GlyMetProGlyGluArgGlylleAlaGlyThrProG 771
QY      1517  CCCAATACCGAGGACACCTGAGGTGCTAGCAGTGCAGGACACCTGATGACACGAGC--- 1573
Db      771  ly-Pro---LysGlyAspArgGlyClylleGlyGlyAlaGlyThrAlaGly 789
QY      1574  -----TTCTCCAGCGCCCTACAGCTGCAGCTGCAGCTCCCTCCCTAATGGACAC 1618
Db      790  AsnAspGlyAlaGlyLeuProGlyProGlyLeu---GlyProGlyProAlaGlyLeu 808
QY      1619  GTGGGTGCTGGAGCAGT-----GGCCTGCTCCACCTCCA----- 1654
Db      809  LeuGlyGlyLysGlyGluProGlyProArgGlyLeuValGlyProGlySerArgGly 828
QY      1655  ---CCCGCGCTCTCGGGGCTCTGCTGTGATCTCCGTACGTGTG----- 1699
Db      829  AsnProGlySerArgGlyGluAsnGlyProThrGlyAlaValGlyPheAlaGlyProGln 848
QY      1700  -----GTGGTGGGTGAGCCACCGAGCCAGCGGTGCTTCGGGC 1738
Db      849  GlySerAspGlyGlnProGlyValLysGlyGluProGlyGlu-ProGlyLysGlyA 868
QY      1739  CGGGCATCTGCGGACCTGCG-----CATCTGGATAGTGCCTTCCTGCTG 1786
Db      868  spAlaGlySerProGlyProGlnGlyLeuAlaGlySerProGly----- 882
QY      1787  TCCAGGTGGCCCATCCCTGTTATGGGTCCAT-----TGTCAGCTCAGC 1834
Db      883  --ProHisGlyProAsnGlyValProGlyLeuLysGlyGlyArgGlyThrGlnGlyPro 902
QY      1835  CAGTCTGCTACTGCTATATGCTCTGCGCGAGCGCTGGTCTGGTCGCCATTACTTT 1894
Db      902  ro-----GlyAlaThrGlyPheProGlySerAla----- 911
QY      1895  GCTACACAGTAGTATTGTGACAGAGCGACTTGGCCAAATACTACGGTAGAAAACTTC 1954
Db      911  ----- 911
QY      1955  AGCATATTGGTGGAGGCGCTGCTCCTACCTGGCTCCAGCTCCCGCTCTCTGTAGCCGC 2014
Db      912  -----GlyArgValGlyProProGlyProAlaGlyAla-ProGlyProAlaGlyPro 928
QY      2015  ATGGGGTGGCGGCG-----TGCGCGCCA 2038
Db      929  LeuGlyGluProGlyLysGlyGluGlyProPro 938

RESULT 26
CA21_CHICK
ID  CA21_CHICK STANDARD; PRT; 1362 AA.
AC  P02467; Q90795; P87492; Q90758; Q92014; P87491; Q90792;
DT  21-JUL-1986 (Rel. 01, Created)
DI  15-DEC-1998 (Rel. 37, Last sequence update)
DE  16-OCT-2001 (Rel. 40, Last annotation update)
GN  Collagen alpha 2(I) chain precursor (Fragments).
OS  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus
OX  NCBI_TaxID=9031;
RN  [1]
RP  SEQUENCE OF 1-245; 262-448 AND 466-1362 FROM N.A.
RX  MEDLINE=86185168; PubMed=3868961;
RA  Boedtker H., Finer M., Aho S.;

```

```

RT  "The structure of the chicken alpha 2 collagen gene.";
RL  Ann. N.Y. Acad. Sci. 460:85-116(1985).
RN  [2]
RP  SEQUENCE OF 1-89 FROM N.A.
RX  MEDLINE=83246518; PubMed=6135195;
RA  Tate V.E., Finer M.H., Boedtker H., Doty P.;
RT  "Chick pro alpha 2.(I) collagen gene: exon location and coding
RL  potential for the prepropeptide.";
RN  Nucleic Acids Res. 11:91-104(1983).
RN  [3]
RP  SEQUENCE OF 1-14 FROM N.A.
RX  MEDLINE=82060240; PubMed=6946474;
RA  Vogeli G., Ohkubo H., Sobel M.E., Yamada Y., Pastan I.,
RN  de Crombrughe B.;
RT  "Structure of the promoter for chicken alpha 2 type I collagen gene.";
RL  Proc. Natl. Acad. Sci. U.S.A. 78:5334-5338(1981).
RN  [4]
RP  SEQUENCE OF 1-33 FROM N.A.
RX  MEDLINE=84297217; PubMed=6473103;
RA  Aho S., Tate V.E., Boedtker H.;
RT  "Location of the 11 bp exon in the chicken pro alpha 2(I) collagen
RL  gene.";
RN  Nucleic Acids Res. 12:6117-6125(1984).
RN  [5]
RP  SEQUENCE OF 1-79 FROM N.A.
RX  MEDLINE=88056316; PubMed=3678834;
RA  Finer M.H., Boedtker H., Doty P.;
RT  "Construction and characterization of cDNA clones encoding the 5' end
RL  of the chicken pro alpha 1(I) collagen mRNA.";
RN  Gene 56:71-78(1987).
RN  [6]
RP  SEQUENCE OF 78-92.
RC  TISSUE=Skin;
RX  MEDLINE=71115216; PubMed=5544653;
RA  Highberger J.H., Kang A.H., Gross J.;
RT  "Comparative studies on the amino acid sequence of the alpha 2-CB2
RN  peptides from chick and rat skin collagens.";
RN  Biochemistry 10:610-616(1971).
RN  [7]
RP  SEQUENCE OF 74-91; 263-448 AND 1088-1169 FROM N.A.
RX  MEDLINE=82058081; PubMed=6272119;
RA  Wozney J., Hanahan D., Tate V.E., Boedtker H., Doty P.;
RT  "Structure of the pro alpha 2 (I) collagen gene.";
RL  Nature 294:129-135(1981).
RN  [8]
RP  SEQUENCE OF 78-92.
RC  TISSUE=Skin;
RX  MEDLINE=70131186; PubMed=4313735;
RA  Kang A.H., Gross J.;
RT  "Amino acid sequence of cyanogen bromide peptides from the amino-
RN  terminal region of chick skin collagen.";
RL  Biochemistry 9:796-804(1970).
RN  [9]
RP  SEQUENCE OF 78-92 AND 415-448.
RC  TISSUE=Skin;
RX  MEDLINE=69285369; PubMed=5809220;
RA  Kang A.H., Igarashi S., Gross J.;
RT  "Characterization of the cyanogen bromide peptides from the alpha-2
RN  chain of chick skin collagen.";
RL  Biochemistry 8:3200-3204(1969).
RN  [10]
RP  SEQUENCE OF 78-92 AND 415-448.
RC  TISSUE=Bone;
RX  MEDLINE=69206882; PubMed=5785233;
RA  Lane J.M., Miller E.J.;
RT  "Isolation and characterization of the peptides derived from the
RN  alpha 2 chain of chick bone collagen after cyanogen bromide
RT  cleavage.";
RL  Biochemistry 8:2134-2139(1969).
RN  [11]
RP  SEQUENCE OF 566-587 FROM N.A.
RX  MEDLINE=79074829; PubMed=364479;
RA  Lehrach H., Frischauf A.-M., Hanahan D., Wozney J., Fuller F.,

```

Crkvenjakov R., Boedtker H., Doty P.;
 "Construction and characterization of a 2.5-kilobase procollagen
 clone";
 Proc. Natl. Acad. Sci. U.S.A. 75:5417-5421(1978).
 [12]
 SEQUENCE OF 902-1362 FROM N.A.
 MEDLINE-81160715; PubMed-6927845;
 Fuller F., Boedtker H.;
 "Sequence determination and analysis of the 3' region of chicken pro-
 alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
 including the carboxy-terminal propeptide sequences.";
 Biochemistry 20:996-1006(1981).
 [13]
 SEQUENCE OF 998-1169 AND 1234-1362 FROM N.A.
 MEDLINE-81264246; PubMed-6267043;
 Dickson L.A., Nimniy Y., Bernard M.P., Pesciotta D.M., Parsons J.,
 Green G., Eikenberry E.F., de Crombrughe B., Vogeli G., Pastan I.,
 Fietzek P.P., Olsen B.R.;
 "The exon/intron structure of the 3'-region of the pro alpha 2(I)
 collagen gene.";
 J. Biol. Chem. 256:8407-8415(1981).
 [14]
 SEQUENCE OF 932-954 AND 968-980 FROM N.A.
 MEDLINE-81064671; PubMed-6159882;
 Avvedimento V.E., Vogeli G., Yamada Y., Maizel J.V. Jr., Pastan I.,
 de Crombrughe B.;
 "Correlation between splicing sites within an intron and their
 sequence complementarity with UI RNA.";
 Cell 21:689-696(1980).
 [15]
 SEQUENCE OF 126-161; 467-517 AND 926-954 FROM N.A.
 MEDLINE-81112157; PubMed-7460017;
 Yamada Y., Avvedimento V.E., Mudryj M., Ohkubo H., Vogeli G.,
 Iranli M., Pastan I., de Crombrughe B.;
 "The collagen gene: evidence for its evolutionary assembly by
 amplification of a DNA segment containing an exon of 54 bp";
 Cell 22:887-892(1980).
 C-1 FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 (FIBRILLAR FORMING COLLAGEN).
 C-1 SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 C-1 TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 HYDROXYAPATITE.
 C-1 PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

 ENBL; M25963; AAA69960.1; -
 ENBL; M25956; AAA69960.1; JOINED.
 ENBL; M25959; AAA69960.1; JOINED.
 ENBL; M25961; AAA69960.1; JOINED.
 ENBL; M25962; AAA69960.1; JOINED.
 ENBL; M25965; AAA69961.1; -
 ENBL; M25964; AAA69961.1; JOINED.
 ENBL; M25984; AAA69962.1; -
 ENBL; M25957; AAA69962.1; JOINED.
 ENBL; M25966; AAA69962.1; JOINED.
 ENBL; M25967; AAA69962.1; JOINED.
 ENBL; M25969; AAA69962.1; JOINED.
 ENBL; M25970; AAA69962.1; JOINED.
 ENBL; M25971; AAA69962.1; JOINED.
 ENBL; M25972; AAA69962.1; JOINED.
 ENBL; M25973; AAA69962.1; JOINED.
 ENBL; M25974; AAA69962.1; JOINED.
 ENBL; M25976; AAA69962.1; JOINED.
 ENBL; M25977; AAA69962.1; JOINED.


```

QY 1795 CACCTGGACACGACGAGGAC---TATCCAGGATGGCGAGGTCCAGGACGATGCCCGC 1739
Db 136 roProGlyLysAlaGlyGluAspGlyHisProGly-----LysProGlyArg---ProG 153
QY 1738 GC-----CCGGAACAC 1727
Db 153 lyGluArgGlyValAlaGlyProGlyGlyAlaArgGlyPheProGlyThrProGlyPro- 172
QY 1726 CCTGGCTCGTGGCTACCCACCCACACGATCGGACATCAGCAGCACA-----1673
Db 173 ProGlyPheLysGlyIleArgGlyHisAsnGlyLeuAspGlyLeuThrGlyGlnProGly 192
QY 1673 -----1673
Db 193 AlaProGlyThrLysGlyGluProGlyAlaProGlyGluAsnGlyThrProGlyGlnPro 212
QY 1672 -----GGCCCGCAGAGCGC 1658
Db 213 GlyAlaArgGlyLeuProGlyGluArgGlyArgIleGlyAlaProGlyProAla-GlyAl 232
QY 1657 GGTTGGA-----GGTGGAGACGAGCCACTGCCTCCAGCA-----1624
Db 232 aaArgGlySerAspGlySerAlaGlyProThrGlyProAla*****252
QY 1624 -----1624
Db 252 *****GlyGluIleGlyProAlaGlyAsnGluG 272
QY 1623 -CCACGTGTCCATTAGG-----1606
Db 272 yProThrGlyProAlaGlyProArgGlyGluIleGlyLeuProGlySerSerGlyProVa 292
QY 1605 -----AAGGAGCTCCAGGCTTAGGGCTGGCAGGAAGCTGGTCAATCAG 1562
Db 292 lGlyProProGlyAsnProGlyAlaAsnGlyLeu---ProGlyAlaLys-GlyAlaAlaG 311
QY 1561 GCTGTCTCTACTGCTAGCACCTCCAGTGTCCCTCGTATTGGGAGGAACACCTGCTT 1502
Db 311 lyLeuPro-----GlyVal---AlaGlyAlaProGlyL 321
QY 1501 CTCCCGTGTGTAGAGG-----AGCCAGTGTGTAGGCGAGGATGTGCAG 1457
Db 321 euProGlyProArgGlyIleProGlyProGlyProAlaGlyProSerGlyAlaArgG 341
QY 1456 GGCTGAGAGGTGAACCCGGTGAGGCGGCTGAAGCTGTCAACAGGCGCAC-----ACT 1403
Db 341 lyLeu-ValGlyGluProGlyProAlaGlyAlaLysGlyGlySerGlyAsnLysGlyGlu 360
QY 1402 GTGGACAGCATGTGGCACCGCACCGACAGGAAAGCTGCCACTGGCCAAATAGAC 1343
Db 361 ProGlyAlaAlaGlyProGlyProGlyProGlyProSerGlyGlyGluGlyLysArgGly 380
QY 1342 TGCTGAGTGCCTGAATC-----GCTGCACCGACCGGTCCA-----TGA 1305
Db 381 ---SerAsnGlyGluProGlySerAlaGlyProGlyProGlyProAlaGlyLeuArgGlyGlu 399
QY 1304 CCAGAGAGAGA-----CCAGGAGATGGCGCATCTGCAGGAACA-----1266
Db 400 ProGlySerArgGlyLeuProGlyAlaAspGlyArgAlaGlyValMetGlyProAlaGly 419
QY 1266 -----1266
Db 420 AsnArgGlyAlaSerGlyProValGlyAlaLysGlyProAsnGlyAspAlaGlyArgPro 439
QY 1265 -----GCCCGCAGGTGCCATCCGACCGCTTTCATCATGTGTCTCC-----1224
Db 440 GlyGluProGlyLeuMetGlyProArg*****459
QY 1223 -----GGCCCTCGGTGC 1212
Db 460 *****GlyPheProGlyAlaAspGlyArgValGlyProIleGly 479
QY 1211 CCGGCTCAGCTCTGGGCGACCGCTCGGTACAGCC-----CCTCGCCCGACGAATCCGTGT 1158

```

```

Db 480 ProAlaGlyAsnArgGlyGluProGlyAsnIleGlyPheProGlyProLysGlyPro--- 498
QY 1157 AAACACGCTGAAGTTCATGAGTGCATCCACTGCACAGCTCACAGGTCACCCAGGACCGGC 1098
Db 499 ---ThrGlyGluProGlyLysProGly 506
QY 1098 -----1098
Db 507 GluLysGlyAsnValGlyLeuAlaGlyProArgGlyAlaProGlyProGluGlyAsnAsn 526
QY 1097 -----CCAGGTCGGGGCCATCGGCGGAGCAGCACA-----1071
Db 527 GlyAlaGlnGlyProGlyValThrGlyAsnGlnGlyAlaLysGlyGluThrGlyPro 546
QY 1070 -----GCTGTGTGACCGGCGGAGCAGCGG 1047
Db 547 AlaGlyProGlyPheGlnGlyLeuProGlyProSerGlyProAlaGlyGluAlaGly 566
QY 1046 CCCCCAGGTTCC-----GGAAG-----CCAAGCGGCGCC---1017
Db 567 LysProGlyGluArgGlyLeuHisGlyGluPheGlyValProGlyProAlaGlyProArg 586
QY 1016 -----GGCATG---GACAGCAGTGGGCGGACAGAGGCGGCGCCACAGCCCTTCTG 969
Db 587 GlyGluArgGlyLeuProGlyGluSerGlyAlaVal-----GlyProAlaGlyProIle 604
QY 968 CTGGCTCGTGGGCGCCAGCGCTGCTCTCAGCCACCCAGCAGTGTGTGCTCTACGCGAGG 909
Db 605 GlySerArg---GlyProSerGlyProGlyProGlyProAsp-----616
QY 908 TGAGAAATCAGGTGTAGCAGCGCCAAAGAGGCTCTCTCTGTTGGTGGCCAGGT---AGG 852
Db 617 ---GlyAsnLysGlyGluProGlyAsnValGly---ProAlaGlyAlaProGlyProAla 634
QY 851 GGGCAGGCGCTGTGTCTCCAGTCAAGTGGCAGGAGGAGTGTGCTGAGGCGGCGGCGCCCA 792
Db 635 GlyProGly---GlyIleProGlyGluArgGlyValAlaGlyValProGlyGly-----651
QY 791 GACTGATCATGAAGCATAGACAGATAGGCT-----GGCGACAGTGGT---747
Db 652 -----LysGlyGluLysGlyAlaProGlyLeuArgGlyAspThrGlyAlaThr 667
QY 746 -----CCGGGTCC-----GGA 735
Db 668 GlyArgaspGlyAlaArgGlyLeuProGlyAlaIleGlyAlaProGlyProAlaGlyGly 687
QY 734 AGAGTTCAGAGA-----GCAGGCGCTCCAGTGGAGTGAAGC 699
Db 688 AlaGlyAspArgGlyGlyGlyGlyProAlaGlyProAlaGlyProAlaGlyAlaArgGly 707
QY 698 ACACCTGGCCACAAAGTCCAGCAGCCCGCCAGGATGAGCAGTGCAGCTCCAGG 639
Db 708 IleProGlyGluArgGly-----GluProGlyProValGlyProSerGlyPhe 723
QY 638 GCTTGGGATCGGCGCAGCAGCCCTGCTAGCCAGCCGG-----600
Db 724 AlaGlyProGlyAlaAlaGlyGlnProGlyAlaLysGlyGluArgGlyProLysGly 743
QY 599 CCCTTGGGATGAGAAAGGCTCAGCAGGATGCCAGGACAGTGCAGGCGGCGGCGGCGG 540
Db 744 ProLysGly---GluThrGlyProThrGly-AlaIleGlyProIleGlyAlaSerGlyPr 762
QY 539 GCGCGCGCGCCATA-----GCGTCCACGCCAGT 513
Db 762 oProGlyProValGlyAlaAlaGlyProAlaGlyProArgGlyAspAlaGlyProGly 782
QY 512 GGTCACTGGCTGAGCTAGGAGCGGCGGACACAGACAGGCGGCGGCGGCGGCGGCGGCGG 453
Db 782 yMetThrGly-PhePro-GlyAlaAlaGlyArgValGlyPro-----Pro 796
QY 452 GCACCATGTCATGAATCTCTCTCTA---CCGCCACTTCCAGCAGCAGGCGGCGGCGCAT 396

```


MM; 267750; --
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF01391; Collagen; 7.
 DR Pfam; PF02210; TSPN; 1.
 DR SMART; SM00282; Lam; 1.
 DR SMART; SM00210; TSPN; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
 KW Polymorphism.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 1516 COLLAGEN ALPHA 1(XVIII) CHAIN.
 FT CHAIN 1334 1516 ENDOSTATIN.
 FT DOMAIN 24 516 NONHELICAL REGION 1 (NC1).
 FT DOMAIN 517 550 TRIPLE-HELICAL REGION 1 (COL1).
 FT DOMAIN 551 560 NONHELICAL REGION 2 (NC2).
 FT DOMAIN 561 640 TRIPLE-HELICAL REGION 2 (COL2).
 FT DOMAIN 641 664 TRIPLE-HELICAL REGION 3 (NC3).
 FT DOMAIN 665 786 TRIPLE-HELICAL REGION 3 (COL3).
 FT DOMAIN 787 809 TRIPLE-HELICAL REGION 4 (NC4).
 FT DOMAIN 810 892 TRIPLE-HELICAL REGION 4 (COL4).
 FT DOMAIN 893 906 NONHELICAL REGION 5 (NC5).
 FT DOMAIN 907 948 TRIPLE-HELICAL REGION 5 (COL5).
 FT DOMAIN 949 961 NONHELICAL REGION 6 (NC6).
 FT DOMAIN 1035 1044 TRIPLE-HELICAL REGION 6 (COL6).
 FT DOMAIN 1045 1077 TRIPLE-HELICAL REGION 7 (NC7).
 FT DOMAIN 1078 1089 TRIPLE-HELICAL REGION 7 (COL7).
 FT DOMAIN 1090 1111 NONHELICAL REGION 8 (NC8).
 FT DOMAIN 1112 1118 TRIPLE-HELICAL REGION 8 (COL8).
 FT DOMAIN 1119 1173 TRIPLE-HELICAL REGION 9 (NC9).
 FT DOMAIN 1174 1186 TRIPLE-HELICAL REGION 9 (COL9).
 FT DOMAIN 1187 1204 TRIPLE-HELICAL REGION 10 (NC10).
 FT DOMAIN 1205 1516 TRIPLE-HELICAL REGION 10 (COL10).
 FT CARBOHYD 68 NONHELICAL REGION 11 (NC11).
 FT CARBOHYD 129 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 164 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 691 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1329 O-LINKED (GALNAC. .).
 FT DISULFID 1366 /FTid=CAR_000150.
 FT DISULFID 1468 BY S-MILARITY.
 FT SITE 1095 CELL ATTACHMENT SITE (POTENTIAL).
 FT VARSPLIC 1 180 MISSING (IN SHORT ISOFORM).
 FT VARSPLIC 181 HTACTLPAPTPSPSLGRPWAPLTGPSVPPSS -> MA
 FT VARIANT 1437 PRCPWPERRLLDVLAPLVLLGVRAASAE (IN
 FT VARIANT 1437 SHORT ISOFORM).
 FT VARIANT 1437 D -> N.
 FT CONFLICT 428 /FTid=VAR_012709.
 FT CONFLICT 841 F -> S (IN REF. 2).
 FT CONFLICT 877 I -> V (IN REF. 2).
 FT CONFLICT 886 V -> L (IN REF. 3).
 FT CONFLICT 912 P -> R (IN REF. 3).
 FT CONFLICT 933 P -> R (IN REF. 3).
 FT CONFLICT 975 R -> L (IN REF. 3).
 FT CONFLICT 1064 P -> L (IN REF. 3).
 FT CONFLICT 1084 A -> P (IN REF. 3).
 FT CONFLICT 1120 L -> K (IN REF. 3).
 FT CONFLICT 1123 P -> A (IN REF. 3).
 FT CONFLICT 1126 P -> A (IN REF. 3).
 FT CONFLICT 1206 G -> PGFP (IN REF. 2).
 FT CONFLICT 1304 R -> GQ (IN REF. 3).
 FT CONFLICT 1314 A -> G (IN REF. 3).
 FT CONFLICT 1323 LR -> CG (IN REF. 3).
 FT CONFLICT 1443 R -> T (IN REF. 4).
 FT CONFLICT 1483 S -> Y (IN REF. 4).
 SQ SEQUENCE 1516 AA; 153840 MW; 3C70F29A4476EE76 CRC64;
 Alignment Scores:
 Pred. No.: 2,29e-07
 Score: 311.50
 Length: 1516
 Matches: 289

Percent Similarity: 29.28% Conservative: 70
 Best Local Similarity: 23.57% Mismatches: 367
 Query Match: 4.85% Indels: 502
 DB: 1 Gaps: 67
 US-09-759-143-110 (1-3410) x CA1H_HUMAN (1-1516)
 QY 93 CAGCAGCAGGCTTTCAGCATGGCTGAGAAAGCTGACCGGACCAAGGGCTGGCAGAAA 152
 Db GlnThrThrValAlaSerLeuGlnThrLeuProGlySerAspSerValSerThr 501
 QY 153 TGGCGCGCTGGCTGATTCTTAGCGAGTTGGCGGAGCAGGAGGAGCGCCAGCTTCT 212
 Db 502 TrpAsp-----GlySerValArgThrProGlyGlyArgValLysGluGly 516
 QY 213 GGAGCAGAGCCGAGCAGAGCAGTTCCTGAGTGCCTGAACGCCGCCCTGACCCCTACCCG 272
 Db 517 GlyLeuLysGlyGlnLysGlyGluProGlyValProGlyProGlyArgAlaGlyPro 536
 QY 273 CCTGGC-----CCACTATGG 287
 Db 537 ProGlySerProCysLeuProGlyProGlyLeuProCysProValSerProLeu-Gl 556
 QY 288 T-----CCAGAGGCTGTGGGTGAGCGCGCTGC 314
 Db 556 yProAlaGlyProAlaLeuGlnThrValProGlyProGlyProGlyProGlyProGly 576
 QY 315 T-----GCGGCACCGGAAAGCCAGCTCTTGTGGTCAACC 350
 Db 576 yArgAspGlyThrProGlyArgAspGlyGluProGlyAspProGlyGluAspGlyLysPr 596
 QY 351 TGCTAACCTTTGGCTGAGGTGTGTTGGCGGAGGAGCATCACATATGTGCGCGCTGTC 410
 Db 596 oGlyAspThrGlyProGlyGlyPheProGly-----ThrPr 608
 QY 411 TGCTGAAAGTGGGGTAGAGGAGAGTTCATGACCATGGTCTGGGCAATTGGTCCAGTGC 470
 Db 608 oGlyAspValGlyProLysGlyAspLysGlyAspProGlyValGlyGluArgGlyProPr 628
 QY 471 TGGGCTGCTGTGTGCCCTCTAGCTCAGCCAGTACCATGCGCTGGCGTGGAGCGCTATG 530
 Db 628 oGlyProGlnGlyProProGlyPro----- 636
 QY 531 GCCCGCGCGCGCTTCAT-----CTGGGCATGTCTCT 563
 Db 637 -ProGlyProSerPheArgHisAspLysLeuThrPheIleAspMetGluGlySerGlyPh 656
 QY 564 TGGGCATCTCTGAGCCTCTTCTTCATCCCAAGCGCGCTGGCTAGCAGGCGTGTGT 623
 Db 656 eGlyGlyAspLeuGluAlaLeuArgGlyProArgGly-----ph 669
 QY 624 GCCCGATCCCGCGCGCTGGAGCTGGCACTGCTCATCTCTGGCGGT----- 669
 Db 669 eProGlyProGlyProGlyProGlyValProGlyLeuProGlyGluProGlyArgPheGl 689
 QY 670 -----GGGCTGCTGGACTTCTGTGGCCAGGTGCTGCTACTC 707
 Db 689 yValAsnSerSerAspValProGlyProGlyProGlyProGlyProGlyProGlyAr 707
 QY 708 CACTGGAGGCGCTCTCTGACCTCTTCCGGGAGCC-----GG 746
 Db 707 gGluGlyProGlyPheProGlyLeuProGlyProGlyProGlyProGlyProGlyGluGl 727
 QY 747 ACCACTG-----TCGCCAGGCCCTACT 767
 Db 727 yProGlyArgThrGlyGlnLysGlySerLeuGlyGluAlaGlyAlaProGly----- 745
 QY 768 CTGTCTATGCCCTTATGATCATGCTTGGGGCTGCTGGGCTGCTGCTGCTGCCATTG 827
 Db 746 -----HisLysGlySerLysGlyAlaProGly-----ProAla----- 756
 QY 828 ACTGGGACACCAG-----TGCCCTGGCGCCCTA-----CCTGG 860
 Db 828 ACTGGGACACCAG-----TGCCCTGGCGCCCTA-----CCTGG 860

Db 757 -----GlyAlaArgGlyGlySerGlyLeuAlaGlyAlaProGlyProAlaGlyProGly 775
QY 861 GCACCCAGGAGGAGTGCCTCTTTGGCCCTGCTACACCTCATCTCTCCTACCTGCGTAGCAG 920
Db 775 yProProGly-----ProProGlyProProGlyProGlyLeuPro----- 788
QY 921 CCACACTGCTGCT-----GGCTGAGGAGGAGGCTGGGCCCCAC---CGAGCCAGCAG 971
Db 789 -----AlaGlyPheAspAspMetGluGlySerGlyGlyProPheTrpSerThrAlaAr 806
QY 972 AAGGCTGTGGCCCTCTCTGTCGCCCCACTGCTGTCATGCGGGCCCGCTTGCGCTT 1031
Db 806 gSerAlaAspGlyPro-----GlnGlyProProGlyLe 817
QY 1032 TCCG-----GAACCTGGCGCCCTGCTTCCCGCGCTGCACACAGCTGTGCTGCC 1079
Db 817 uProGlyLeuLysGlyAspProGlyValProGlyLeuProGlyAlaLysGly----- 834
QY 1080 GCATGCCCGCCACCTGCGCGGCTCTTCTGGCTGAGCTGTCAGCTGGATGGCACTCA 1139
Db 835 -----GluValGlyAlaAspGly----- 840
QY 1140 TCACCTTCACGCTGTTTACACCGATTCGTGGCGGAGGGCTGTACAGGGCGTGCCTCA 1199
Db 841 -----IleProGlyPheProGly-----LeuProGlyArg---Gly 851
QY 1200 GAGCTGAGCGCGCCAGCGAGGCGCGAGACACTATGATGAAGCGCTTCGGAT---GGCA 1256
Db 851 uGlyIleAlaGlyProGlnGlyProLysGlyAspArgGlySerArgGlyGlyLysGlyAs 871
QY 1257 GCTGGGGCTGTCTCTGTCAGTGCAGCTCCTCCTGGTCTCTCTGTCATGACACGGC 1316
Db 871 pProGlyLysAspGlyValGlyGln-----ProGlyLeuProGlyProProGlyProPr 889
QY 1317 TGGTCAGCGATTCGGCACTCAGCAGCTATTTGGCCAGTGTGCGAGCTTCCCTGTGG 1376
Db 889 oGlyProValTyTrValSerGlu-----GlnAspGlySerVal----- 902
QY 1377 CTGCGGTCACATGCTCTCCACAGTGTGCGCGTGGTGCAGACTTCAGCGCGCTCA 1436
Db 903 -----LeuSerValProGlyProGlu-----GlyArgProGly 913
QY 1437 CCGGTTACCTCTCAGCGCTGCAGATCTCCTCCCTACAC-----AC 1478
Db 913 yPheAlaGlyPheProGlyProAlaGlyProLysGlyAsnLeuGlySerLysGlyLuar 933
QY 1479 TGGCTCTCTCTACACCGGAGAGAGAGTGTCTGCGCCAAATACCGAGGGACACTG 1538
Db 933 gGlySerProGlyProLysGlyGlyLysGlyGluProGlySerIle-PheSerProAspG 953
QY 1539 GAGGTCTAGCAGTGAAGACGCTGATGACACGCTTCTGCGAGGCTTCTGCGAGGCTTGA 1598
Db 953 lGlyAlaLeuGlyProAlaGlnLysGlyAlaLysGlyGluProGlyPheArg---GlyP 972
QY 1599 CTCCCTCTCTATGACACAGTGGTGTGAGGAGGAGTGGCTCTCCACCTCCACCGC 1658
Db 972 roProGlyProTyArgGlyArgProGlyTyArgGlyGlyGlyGlyGlyGlyGlyGlyGly 992
QY 1659 CGCTCTGCGGGCTCTGCTGTGATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1718
Db 992 lArgProGlyMetAsn-----GlyLeuLysGlyGlyGlyGlyGlyGlyGlyGlyGly 1008
QY 1719 AGCCAGGCT-GGTTCGGCGCGGCGCATCTCCCTGGACCTCGCCCTCTGATAGTGC 1777
Db 1008 spAlaSerLeuGlyPheGlyMetArgGlyMetProGlyProGlyPro----- 1024
QY 1778 TTCTCTCTCTCCAGTGGCGCC---ATCCCTGTTTATGGGCTCCATGTCAGCTCAGC 1834
Db 1025 --ProGlyProProGlyProGlyThrProValTyAspSerAsn----- 1039
QY 1835 CAGTCTGTCTGCTATATGCTGTCTGCGCGAGCGCTGGTCTGCTGCTGCTGCTGCTGCT 1894
Db 1040 -----ValPheAlaGluSerArgProGlyPro----- 1049

QY 1895 GCTACACAGGTAGTATTTGCAAGAGCGACTTGGCCAAATACTACTACGCTAGAAAACCTTC 1954
Db 1049 ----- 1049
QY 1955 AGCACATTGGGTGGAGGCTCCCTCAGTGGTCCAGCTCCCTCCCTCCT----- 2005
Db 1050 -----ProGlyLeuPro-GlyAsnGlnGlyProGlyProGlyGlyAla 1064
QY 2006 -----GTTAGCCCATGGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCTGCTGCTGCTG 2056
Db 1065 LysGlyGluValGlyProGlyProGlyProGlyGlnPhePro-PheAspPheLeu----- 1082
QY 2057 GTAATGTGGCTCTCTGCTGCCACCTCTGCTGAGTGGTGGTGGTGGTGGTGGTGGG 2116
Db 1083 -----GlnLeuGluAlaGluMetLysGly----- 1090
QY 2117 CTGGGGGCTCCCTCTCTCTCTCCAGTCTCTAGGGCTGCTGAGTGGAGGCTTCCAA 2176
Db 1091 -----GluLy 1092
QY 2177 GGGGTTTTCAGTCTGGATTATACAGGAGGCCAGAGGCTCCATGCTGCTGCTGCTGCTGCT 2236
Db 1092 sGly-----AspArgGlyAspAlaGlyGlnLysGlyGluArgG 1105
QY 2237 GAGCTCTGACAGTGGATTACCCAGGCTAGGGTTAAACAGCTAGCTCCTAGTTGAGACAC 2296
Db 1105 yGluProGlyGlyGly----- 1110
QY 2297 ACCTAGAGAGGGTTTGGGAGCTGATTAACACTCAGTCACTGCTGCTTCCATCTCTAAG 2356
Db 1111 -----GlyPhePheGlySerLeuProGlyProGlyProPro----- 1124
QY 2357 CCCCTTAACCTGCGAGCTTCGTTTAACTAGTCTCTGCTGGAGTCTTCTAGGATGAACA 2416
Db 1124 ----- 1124
QY 2417 CTCCTCATGGATTGAACATATGACTTATTTGTAGGGAAGAGTCTGCTGAGGGGAACA 2476
Db 1125 -GlyProArgGlyTyProGlyIleProGlyProLysGlyGlySer-IleArgGly----- 1142
QY 2477 CACAAGAACCGTCCCTCAGCCACAGCAGCTGCTTTTGTGCTGATCCACCCCTCTT 2536
Db 1143 -----GlnProGlyProGlyProGln-----GlyProPro----- 1153
QY 2537 ACCTTTTATCAGGATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2596
Db 1153 ----- 1153
QY 2597 TAAATATTAACTTATTATTATTAACAAGTAGAAGGAATCCATTGCTAGCTTTCTGTG 2656
Db 1153 ----- 1153
QY 2657 TTGCTGTCTAATATTGCTGAGTGGGATCCCAACAATCAGGTCC-----C 2706
Db 1154 --GlyIleGlyTyGluGlyArgGlnGlyProGlyProGlyProGlyProGlyProGlyProp 1173
QY 2707 CTGAGATAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2759
Db 1173 roSerPheProGly-----ProHisArgGlnThrIleSerValProGlyProGly 1190
QY 2760 -----CTGCGGCG 2798
Db 1190 lyProProGlyProProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGly 1210
QY 2799 --TCTACTCATCCCAATGATAATTCCAATGCTGTTTACCAAGTGTAGGTGTGTGAAG 2856
Db 1210 rpAlaThrArg-----GlnAlaMetLeuGlyGlnValHisGluValProG 1225
QY 2857 AAGTAGAGGGTGG----- 2870
Db 1225 lu-----GlyTrpLeuIlePheValAlaGluGlnGluLeuTyTrValArgValGlnA 1243

SEQUENCE OF 95-415, AND REVISION

RC TISSUE=Skin;
RC MEDLINE=76091874; PubMed=173531;
RA Fietzek P.P., Rexrodt F.W.;
RT The covalent structure of collagen. The amino-acid sequence of
RT alpha2-CB4 from calf-skin collagen.;"
Eur. J. Biochem. 59:113-118(1975).
RN [4]
RP SEQUENCE OF 416-445.
RC TISSUE=Skin;
RC MEDLINE=75008198; PubMed=4412529;
RX Fietzek P.P., Furtmayr H., Kuehn K.;
RA Comparative sequence studies on alpha2-CB2 from calf, human, rabbit
RT and pig-skin collagen.;"
Eur. J. Biochem. 47:257-261(1974).
RN [5]
RP SEQUENCE OF 446-481.
RC TISSUE=Skin;
RC MEDLINE=75059250; PubMed=4435743;
RX Fietzek P.P., Kuehn K.;
RA The covalent structure of collagen: amino acid sequence of the N-
RT terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5
RT from calf skin collagen.;"
Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).
RL HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 355:647-650(1974).
CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC -!- PM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB008683; BAA25171.1; -.
DR PIR; A02866; CGB02S.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF01391; Collagen: 18.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Collagen; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Collagen; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT PROPEP 25 79 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 80 1100 COLLAGEN ALPHA 2(I) CHAIN.
FT PROPEP 1101 1364 CARBOXYL-TERMINAL PROPEPTIDE.
FT MOD_RES 80 80 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 84 84 CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT INVOLVED IN CROSS-LINKING (PROBABLE).
FT MOD_RES 175 175 HYDROXYLATION (PROBABLE).
FT MOD_RES 196 196 HYDROXYLATION (PROBABLE).
FT MOD_RES 262 262 HYDROXYLATION (PROBABLE).
FT MOD_RES 307 307 HYDROXYLATION (PROBABLE).
FT MOD_RES 352 352 HYDROXYLATION (PROBABLE).
FT CONFLICT 157 157 V -> P (IN REF. 3).
FT CONFLICT 187 187 K -> T (IN REF. 3).
FT CONFLICT 211 211 T -> K (IN REF. 3).
FT CONFLICT 298 300 PGA -> ASP (IN REF. 3).
FT CONFLICT 423 424 AT -> TA (IN REF. 4).
SQ SEQUENCE 1364 AA; 129064 MW; 5593F4D6B9ED119A CRC64;

Alignment Scores:
Pred. No.: 2,99e-07 Length: 1364

Score:	309.00	Matches:	251
Percent Similarity:	30.97%	Conservative:	41
Best Local Similarity:	26.62%	Mismatches:	321
Query Match:	4.96%	Indels:	330
DB:	1	Gaps:	49

US-09-759-143-110 (1-3410) x CA21_BOVIN (1-1364)	
QY 2173 GAAGGCTCCAGTCAGGAGCCCTAGA-----GACTGGGAGAGAGG--- 2132	
DB 23 GlnSerLeuGlnGluAlaThrAlaArgLysGlyProSerGlyAspArgGlyProArgGly 42	
QY 2131 ---AGAGGAGCGCCCGCCCGCCAGCTGTGAGCTACGACCTCAGCAGCAGCGGTGGC 2075	
DB 43 GluArgGlyProGlyProPro----- 50	
QY 2074 AGCAGAGCCACATTACTTTGGCAGCAACAGAACTGGCGCCAGCCCGCCAGCCCAT 2015	
DB 50 ----- 50	
QY 2014 GGGGTAACAGGAGCGGGAGCTGGGA-----CCAGTGAGGAGGCGCT--- 1970	
DB 51 -----GlyArgAspGlyAspGlyProGlyProGlyProGlyProGlyProPro 68	
QY 1969 ---CCACCCCATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAAG----- 1924	
DB 69 GlyProGlyLeuGlyGlyAsnPheAla-AlaGlnPheAspAlaLysGlyGlyGlyPr 88	
QY 1923 -----TCGCTCTTGTCAAACTACTACCTGTGTAGCAAGTAATGGCGACAGACC 1874	
DB 88 oGlyProMetGlyLeuMetGlyProArgGlyProGlyProGlyProGlyProGlyProGly 108	
QY 1873 CA-----GGCCTGCGGAGACACCATATAGCAGTGTAGCAGTGTGCTGAGCTG 1826	
DB 108 toGlnGlyPheGlnGlyProGlyProGlyProGlyProGlyProGlyProGlyProAlaG 128	
QY 1825 GACATGAGCGCCATAACAGGATCGGGCCACCTGGGACAGCAGGA----- 1779	
DB 128 lYAlaArgGlyPro-----ProGlyProProGlyLysAlaGlyLysAlaGlyHisP 145	
QY 1778 -----AGGCACT 1772	
DB 145 roGlyLysProGlyArgProGlyGlyArgGlyValValGlyProGlnGlyAlaArgGlyP 165	
QY 1771 ATCCAGGATGCGAGGT---CCAGCAGATGCCCGCGCCGGAACACCCCTGGCTCGGT 1715	
DB 165 heProGlyThrProGlyLeuProGlyGlyPheLysGlyLysAlaArgGlyHisGlnGlyLeuAsp 184	
QY 1714 GGGCTCACC-----CAC 1703	
DB 185 GlyLeuLysGlyGlnProGlyAlaProGlyValLysGlyGlyProGlyAlaProGlyGlu 204	
QY 1702 CACCACAGCTAGGAGACATCACAGCAGA----- 1673	
DB 205 AsnGlyThrProGlyGlnThrGlyAlaArgGlyLeuProGlyGlyGlyArgValGly 224	
QY 1672 -----GGCCCGCAGAGCGGGGTGGAGGTGGGAGAGGCCACTGGCTCCAGCAGCCAC 1619	
DB 225 AlaProGlyProAla-GlyAlaArgGlySerAspGlySerValGlyProValGlyProAl 244	
QY 1618 GTGCTCATAGGAG-----GGAGCTCCAGC----- 1591	
DB 244 aGlyProIleGlySerAlaGlyProGlyPheProGlyAlaProGlyProLysGlyG 264	
QY 1590 -TTAGGGCT-----GGCAGGAAGCT 1571	
DB 264 uLeuGlyProValGlyAsnProGlyProAlaGlyProAlaGlyProArgGlyGluValG 284	
QY 1570 GTCTATCAGGTGCTCCTCAGCTAGCACCTCCAGTGTCCCTCGGTATT----- 1521	
DB 284 yLeuProGlyLeuSerGlyProValGlyProGlyAsnPro-GlyAlaAsnGlyLeup 304	
QY 1520 -----TGGGCAGGAACACCTGCTCTCCCGTG 1493	

Db	304	roGlyAlaLysGlyAlaAlaGlyLeuProGlyValAlaGlyAlaProGlyLeuProGlyP 324
QY	1492	GTAGAGG-----AGCCAGGTGTAGGCGCAGGATCTCAGGCGTGTGAGAGGT----- 1445
Db	324	roArgGlyIleProGlyProValGlyAlaAlaGlyAlaThrGlyAlaArg-GlyLeuVal 343
QY	1444	---GAACCCGGTGAGGCGGCTGAAGCTGTACACCGGCAC-----ACT 1403
Db	344	GlyGluProGlyProAlaGlySerLysGlyGluSerGlyAsnLysGlyGluProGlyAla 363
QY	1402	GTGGGACAGCATGTGGCAGCCGACCCAGGAAAGTGGCCACACTGGCCAAATAGAC 1343
Db	364	ValGlyGlnProGlyProGlyProSerGlyGluGluGlyLysArgGly----- 380
QY	1342	TGCTCAGTGTCCGAATCGCTGCACCGCGGTCCATGACGAGAGAAAGCAGGAGAT 1283
Db	381	---SerThrGlyGlu-IleGlyProAlaGlyPro-----ProGlyProP 394
QY	1282	GGCGCACTGCAGAAC-----GCCCCAGGTGCCCATCCGAAACCCCTTCATCATAGTG 1229
Db	394	roGlyLeuArgGlyAsnProGlySerArgGlyLeuProGlyAlaAlaAspGlyArgAlaGly 414
QY	1228	TCTCCGGGCT-----CGGTGCCGCTCAGCTCAGTCTGGGCGACCCCTG 1187
Db	414	aMetGlyProAlaGlySerArgGlyAlaThrGlyProAlaGlyValArgGlyProAsnG 434
QY	1186	GTACAGCCCTCCGCCACCAAAATCCGTGTAAACAGCGTGAAGGTCTAGTCCCATCA 1127
Db	434	lyAspSerGlyArgProGlyGluProGlyLeuMetGlyProArgGlyPheProGlySerP 454
QY	1126	GCTGCACAGCTCAGCCACAGAGCGCGGAGGTGCGGGCATCGGC----- 1077
Db	454	roGlyAsnIleGlyProAlaGlyLysGluGlyProValGlyLeuProGlyIleAspGlyA 474
QY	1076	-----AGCACAGCTGTGTGAGCGGGAAGCAGCGGCCCGCCAGGTTCGGAA 1031
Db	474	rgProGlyProIleGlyProAlaGlyAlaArgGlyGluProGlyAsnIleGlyPheProG 494
QY	1030	AGCAAGCGGCGCGCATGGAGCAGTGGGCGCACAAAGAGGGGCGCCAGACCCCTTC 971
Db	494	lyProLysGlyProSerGlyAspProGlyLysAlaGlyGluLysGlyHisAlaGlyLeuA 514
QY	970	TGCTGCTGGTGGCGCCAGCCCTCCCTCCTCAG-----CCAC 932
Db	514	laGlyAlaArgGlyAlaProGly-----ProAspGlyAsnAsnGlyAlaGlnGlyProP 532
QY	931	CAGCAGTGTGGCTGCTACGCGAGTGTAGGAGTGTAGGCGTGTAGGCGCCAAAGAGGCACT- 873
Db	532	roGlyLeuGln-----GlyValGlnGlyGlyLysGlyGluGlnGlyProA 547
QY	872	-----CCTCCTGGGTGCCAGGTAGGGGCCAGGCACTGGTGTCCCACTCAATGGCAGG 818
Db	547	laGlyProProGlyPheGlnGly-----LeuProGlyProAlaGlyT 561
QY	817	CAGGAGTAGCCCGCAGGAGCCCGCCAGACTGATCATGAAGGCATACAGAGTAGGCGCTG 758
Db	561	hrAlaGlyGluAlaGlyLysProGlyGlu-----ArgGlyIleProG 575
QY	757	GCACAGTGTGT---CCGGGTCCC-----GGAAGAGTGTAGAGCAGGCGCCCTCAGTGG 707
Db	575	lyGluPheGlyLeuProGlyProAlaGlyAlaArgGlyGluArg---GlyProProGlyG 594
QY	706	AGTGAAGACACCTGGCCACAGAAAGTCCAGCAGCCCGCCAGGATGAGCAGTCCAG 647
Db	594	luSerGlyAlaAlaGlyProThrGlyProIleGlySerArgGlyProSerGlyProProG 614
QY	646	CTCA-----GGGCGCTGGGATCCGGGCACAGCAG 617
Db	614	lyProAspGlyAsnLysGlyGluProGlyValGlyAla-----ProGlyThrAlaG 632
QY	616	CCCTGTAGCAGCGCGCCCTTGGGATGAGAAAGAGCTCAGCAGGATGCCAAGGACAG 557

Best Local Similarity: 25.67% Mismatches: 399			
Query Match: 4.80% Indels: 401			
DB: 1 Caps: 65			
US-09-759-143-110 (1-3410) x CAL33_BOVIN (1-1049)			
QY	26	CCGGGTGACAGCGCGCGCTCGGCCAGGATCTGAGTGATGAGAGCGTGTCCCACTGAGG	85
DB	134	ProGlyGluAspGlyAlaProGlyPro-----MetGlyProArg-GI	147
QY	86	TGCCCCACAGCAGCAGGTGTTG-----	107
DB	147	yAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArgGlyAsnAs	167
QY	108	----ACCATGGCTGAGAGCTGGACCGCACCAAGGCTGGCAGAAATGGGCGCTGG	163
DB	167	pGlyAlaArgGlySerAspGlyGlnProGlyProProGly-----ProProGI	183
QY	164	C-----TGATTCCTAGGAGCTGGCGGCAAGGA-----GGAGAGC	202
DB	183	yThrAlaGlyPheProGlySerProGlyAlaLysGlyGluValGlyProAlaGlySerPr	203
QY	203	CGCAGCTTCTGAGCAGAGCCGAGACGAGCAGTTCGAGTGCCTGACGGGCCCTGA	262
DB	203	oGlySerSerGlyAlaProGlyGlnArgGlyGluProGlyProGlnGlyHisAlaGlyAl	223
QY	263	GCCTACCGCTGGCCCACTATGCT-----CCAGAGCTGTGGTGAG-----C	307
DB	223	aProGlyProProGlyPro-ProGlySerAspGlySerProGlyGlyLysGlyGluMetG	243
QY	308	CGCTGCTCGGCACCGGAAAGCCAGCTCTT-----GCTGGTCAACCTGCTTAAC	358
DB	243	lyProAlaGlyIleProGlyAlaProGlyLeuLeuGlyAlaArgGlyProProGlyProp	263
QY	359	TTTGGCTGGAGTGTGTTGGCCGAGCATCACTATGTGCCCTCTGCTGCTGSA	418
DB	263	roGlyThrAsnGlyValProGlyG-nArg-----GlyAlaAlaGlyG	277
QY	419	GTGGGGGTAGA---GGAGAGTTTCATGACCATGGT---GCTGGCATTTGGTCCAGTCTG	472
DB	277	luProGlyLysAsnGlyAlaLysGlyAspProGlyProArgGlyGluArgGlyGluAlaG	297
QY	473	GG---CCTGGTGTGTCCCGTCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAT	529
DB	297	lySerProGlyIleAla-----	302
QY	530	GCGCGCGCGCGCTTCTATCTGGGCACTGTCTTGGCATCTCTGCTGAGCTCTTTCTC	589
DB	303	----GlyProLysGlyGluAspGlyLysAsp---GlySerProGlyGluProGlyAlaA	320
QY	590	ATCCCAAGGCGCGTGGCTAGCAGGCTGTGCTGCGCGGATCCAGCGCCCTGGAGCTG	649
DB	320	snGlyLeuProGlyAlaAlaGlyGluArgGlyValProGlyPheArgGlyProAlaGlyA	340
QY	650	GCACTGCTCATCTGGG-----	666
DB	340	laAsnGlyLeuProGlyGluLysG-yProProGlyAspArgGlyGlyProGlyProAlaG	360
QY	667	-----CGTGGGCTGCTGAGCTTCTGTGGCCAGGTGTCTTCACTCCACTGGAGCGCTG	721
DB	360	lyProArgGlyValAlaGly-----GluProGlyArgAsnGlyLeuProGlyGlyProG	378
QY	722	CTCTCTGACTCTTCCGAGCCGAGCAGCTGCTGCGCAGGCTACTCTGTCTATGCCCTTC	781
DB	378	lyLeuArgGlyIleProGly---SerProGlyGlyProGly-----	390
QY	782	ATGATCAGTCTTGGGGCTGCTGGCTACCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG	841
DB	391	-----SerAsnGlyLysProGlyProProGlySerGlnGlyGluThrGlyArg----	406
QY	842	GCCTCGCCCTTACCTGGCGACCCAGGA-----GGAGTGCCTCTTTGGC	886
DB	407	--ProGlyPro---ProGlySerProGlyProArgGlyGlnProGlyValMetGlyPheP	425
QY	887	CTGCT-----	891
DB	425	roGlyProLysGlyAsnAspGlyAlaProGlyLysAsnGlyGluArgGlyGlyProGlyG	445
QY	892	-----CACCTCATCTTCTCACCTGGGTAGCAGCCACACTGCT---GGTGGCTGAGGAG	943
DB	445	lyProGlyProGlnGlyProAlaGlyLysAsnGlyGluThrGlyProGlnGlyProProG	465
QY	944	GCAGCGTGGGCCCCCAGCAGCAGCAGAGGCTGTGCGGCCCTCTCTGTGCGCCCCAC	1003
DB	465	lyProThrGlyPro---SerGlyAspLysGlyAspThrGlyProProGlyProGlnGlyL	484
QY	1004	TGCTGTCCATCCGCGCGCGCTTGGCTTTCG---GAACCTGGCGCGCTCTTCCCGCG	1060
DB	484	euGlnGlyLeuProGlyThrSerGlyProProGlyGluAsnGlyLysProGlyGluProG	504
QY	1061	CTGCACAGCTGTGTGCGCATGCCCGCACCTGCGCGCGCTCTCTGTGCTGGTGA	1120
DB	504	ly-----ProLysGlyGluAlaGlyAlaProGlyIleProGly-----	516
QY	1121	TGCAGCTGGATGGCACTCATGACCTTCACTGCTT---TTACACGGATTTCTGGCGGAG	1177
DB	517	-----GlyLysGlyAspSerGlyAlaProGlyGluArgGlyProProGlyAlaG	533
QY	1178	GGCTGTACAGGCGGTGCCAGAGCTGACCGCGGACCGGAGCGGAGACACTATGAT	1237
DB	533	lyGlyProProGly---ProArgGlyGlyAlaGlyProProGlyProGluGly-----G	550
QY	1238	GAAGCGTTCGATGGGAGCGCTGGG-----GCTGTCTTCAGTGCAGTGC	1282
DB	550	lyLysGlyAlaAlaGlyProProGlyProProGlySerAlaGlyThrProGlyLeuGlnG	570
QY	1283	ATCTCCCTGGTCTTCTC-----TCTGTCTATGACCGCTGTGTCAGCATTCGGC	1333
DB	570	lyMetProGlyGluArgGlyGlyProGlyGlyProGlyProLysGlyAspLysGlyGluP	590
QY	1334	ACTCGAGCAGTCTATTTGGCCAGTGTGCGAGCTTCCCTGTGCTGCGGTGCCACATGC	1393
DB	590	roGlySerSer-----GlyValAspGlyAla-----	599
QY	1394	CTGTCCACAGTGTGGCGTGTGACAGTTCAGCGCCCTCACCGGGTTTCACTTCTCA	1453
DB	599	roGlyLysAspGlyProArgGlyProThr-----	609
QY	1454	GCCCTGACAGTCTGCGCTACACACTGGCTCCCTCTACACCGGAGAGACAGTGTTC	1513
DB	609	lyProIleGlyProProGlyProAlaGlyGlnProGlyAspLysGlyGluSerGlyAlaP	629
QY	1514	CTGCC-----CAATACCGAGGAGACACTGGAGGTGCTAGCAGTGGAGCAGC	1561
DB	629	roGlyValProGlyIleAlaGlyProArgGly-----GlyProGlyGluA	644
QY	1562	CTGATGACCACTTCTGCGCAGCGCTTAAG-----CCTGGAGCTCCCTTCCCTAATGGA	1615
DB	644	rgGlyGluGlnGlyPro-ProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGly	663
QY	1616	CACCTGGTGTGAGGAGCTGGCTCTCCCA-----CCTCCACCC	1657
DB	664	GluProGlyAlaLysGlyGluArgGlyAlaProGlyGlyLysGlyGluGlyGlyProPro	683
QY	1658	GCCTCTGCGGGCGCTCTGCCCTGTGATGCTCTCCGTACGTGTGGTGGTGGTGGAGCCACC	1717
DB	684	Gly-AlaAlaGlyProAlaGlyGlySerGlyPro-----AlaGlyProPr	698
QY	1718	G-----AGGCCAGGTGTTCGCGCGCGGCGGATC	1747
DB	698	oGlyProGlnGlyValLysGlyGluArgGlySerPro-GlyGlyProGlyAlaAlaGlyP	718
QY	1748	TGCTGTGACCTCGCCATCTCTGATAGTCCCTTCTCTGCTGCTCCA-----GGTGGCC	1798
DB	718	heProGlyGlyArgGlyPro-----ProGlyProProGlySerAsnGlyAsnP	734

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.

FT SIGNAL 1 24 POTENTIAL.
 FT PROPEP 25 ? AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
 FT CHAIN 2 1096 COLLAGEN ALPHA 2(I) CHAIN.
 FT PROPEP 1097 1356 CARBOXYL-TERMINAL PROPEPTIDE
 (BY SIMILARITY).
 FT CARBOHYD 1257 1257 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1356 AA; 126985 MW; 7BB2F1F80DB10C93 CRC64;

Alignment Scores:

Pred. No.: 3,31e-07 Length: 1356
 Score: 308.00 Matches: 235
 Percent Similarity: 32.39% Conservative: 51
 Best Local Similarity: 26.61% Mismatches: 319
 Query Match: 4.95% Indels: 279
 DB: 1 Caps: 46

US-09-759-143-110 (1-3410) x CA21_ONCMY (1-1356)

2101 CTACGCACCTCAGACGACGAGGTGGCGAGAGAGCCACATTACTTTGGCACAACAGA 2042
 18 LeuLeuAlaSerCysGlnSerGlyGlyLeuLysGlyProArg-----GlyAlaLys 34
 2041 AACTGGCGGCCAGCCCGGACCCATGGGCTAAC---AGGAGCGGGAGCTGGGA--- 1988
 35 GlyProArgGlyAspArgGlyProGlnGlyProAsnGlyArgAspGlyLysAlaGlyLeu 54
 1987 CCCAGTGAGGAGGAGCCCT-----CCACCCCAATGTCTGGAAATTTCT----- 1944
 55 ProGlyIleAlaGlyProGlyProGlyProGlyLeuGlyGlyAsnGlyAlaGlnPhe 74
 1943 -----ACGCTGAGTATTTGGCCAACTCGCTCT 1917
 75 AspGlyGlyLysGlySerAspProGlyProGlyProMetGlyLeuMetGlySerArgGly 94
 1916 TGTCAATACTACCTGTGTAGCAAGTAATGGCGACCAACCA-----GGC 1869
 95 ProAsnGlyProPro-----GlyAlaProGlyProGlnGlyPheThrGly 109
 1868 CTGGCGGACACCATATAGCAGTACAGACTGGTGTAGCTGGCAATGGAGCCCATAA 1809
 110 HisAlaGlyGluProGlyGluProGlyGlnThrGlySerIleGlyAlaArgGlyProThr 129
 1808 ACAGGATGGGCGCCACCTGGGACAGCAGGAGGCAC-----TATCAGAGATGG 1761
 130 GlySerAlaGlyLysProGlyGluAspGlyAsnGlyArgProGlyLysProGlyAsp 149
 1760 CGA---GGTCCAGGACAGATGCCCGCGCCGGAACCCAC-----CCTGGCCTCGGTGG 1713
 150 ArgGlyGlyProGly-ThrGlnGlyAlaArgGlyPheProGlyThrProGlyLeuProGly 169
 1712 GCTACCCACACACAGTACGAGGACATCAGAGCAGA----- 1673
 169 yMetLysGlyHisArgGlyTyAsnGlyLeuAspGlyArgLysGlyGluSerGlyThrAl 189
 1673 ----- 1673
 189 aGlyAlaLysGlyGluThrGlyAlaHisGlyAlaAsnGlySerProGlyProAlaGlySe 209
 1672 -----GGCCCCGAGAGCGCGGTG 1653
 209 rArgGlyLeuAsnGlyGluArgGlyArgAlaGlyProAlaGlyProAla-GlyAlaArg 229
 1652 GAGGTGGGAGGAGCCACTGCTCCAGCACCACCGCTGTCCATTAGG----- 1606
 229 lAlaAspGlySerThrGlyProAlaGlyProAlaGlyProLeuGlyAlaAlaGlyProp 249
 1605 -----AAGGGAGCTCCAGGC----- 1591
 249 roGlyPheProGlyAlaProGlyProLysGlyGluIleGlyAlaGlySerAsnGlyP 269
 1590 -----TTAG 1587

269 rOSerGlyProGlnGlyArgGlyGluProGlyIleAsnGlyAlaValGlyProValG 289
 1586 GGCCT-----GGCAGGAAGCTGTCTCATCAGGTGT 1557
 289 lProValGlyAsnProGlyAsnAsnGlyIleAsnGlyAlaLys-GlyAlaAlaGlyLeu 308
 1556 CCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATTTGGGAGGAGAACACCTGCTTCC 1497
 309 Pro-----GlyVal---AlaGlyAlaProGlyPhePro 318
 1496 GGTGTAGAGGAGGAGCCAGCTGTGTAGG-----GCAGGATCTCAGGCGTG 1452
 319 GlyProArgGlyGlyProGlyProGlnGlySerThrGlyAlaArgGlyLeu 338
 1451 AGAAGGTGAACCC-----GGTGAGGCGGCTGAAGCTGTCAACACG 1410
 339 -GlyGlyAspProGlyProSerGlyGlnLysGlyAspSerGlyAlaLysGlyGluProGly 358
 1409 CCACACT-----GTGGGACAGGATGTGGACCCGACCCACAGGAGAAAGCTGCCACAC 1356
 358 yHisSerGlyValGlnGlyAlaAlaGlyProAlaGlyGluGlyLysArgGlySerThr 378
 1355 TCGCAAAATAGACTCTCGAGTGGCGAATCGCTGCACCCAGCGGTCCATGACAGAGAGA 1296
 378 rGlyGluValGlyAlaThrGlyProAlaGlyLeu-----ArgGly 391
 1295 AGACCAAGGAGATGGCGCACTGCAGGAACAGCCCAAGCTG-----CCCA 1251
 391 yAlaArgGlyGly-AlaGlyThrArgGlyLeuProGlyLeuGluGlyArgGlyProI 411
 1250 TCGCAACGCTTCATCATAGTGTCTCCGGCTCGTCCGCTCGCTCGCTCGCTCGCTCG 1191
 411 lGlyMetProGlyAla-----ArgGlyAlaThrGlyProGlyGlyIleArgGlyAla- 428
 1190 CTGTGTACAGCCCTCGCCACGAAATCGGTAAACAGCGTGAAGCTCATGAGTGCCA 1131
 429 ProGlyAspAlaGlyArgAlaGlyGluSerGlyLeuThrGlyAlaArgGlyLeuProGly 448
 1130 TCCAGCTGCAGCTCAGCCACAGAGCCGCGGAGGTGCGGGCATCGGC----- 1077
 449 AsnSerGlyGlnGlyProGlyLysGlyGluProGlyProGlyAlaAlaGlyLeuAsp 468
 1076 -----AGCACAGCTGTGTGCGCGGAGAGCGCGCCAGGTTC----- 1035
 469 GlyArgThrGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyPhe 488
 1034 GGAAGCAAGCGCGCGCATGGACAGCAGTGGGCGACAGAGGGGGCGGCGGACAGCC 975
 489 ProGlyProLysGlyProGlyGlyGluAlaGlyLysGlyGlyAspLysGlyProThrGly 508
 974 CTTCGTGCTCGTGGTGGGCGCGCGCTG-----CCTCCTCAGCCACCA 930
 509 AlaThrGlyLeuArgGlyGlyProGlyAlaAspGlyAsnGlyAlaProGlyProAla 528
 929 GCAGTGTGCTGTACAGAGTGAAGATGAGGTGAGCAGGAGCAAGAGGAGGAGGAGGAG 870
 529 GlyValVal-----GlyAsnThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 542
 869 CTGGTGTGCGCGCTG-----AGGGGCGCAGGCGCTGTGTCCTCCAGTCAATGG 822
 543 AlaGlyAlaProGlyPheGlnGlyLeuProGlyProAlaGlyProAlaGlyGluAlaGly 562
 821 CAGGAGGAGGAGTGGCCAGGAGCCCAAGACTGATCATGAAGCATGACAGAGAGTAGG 762
 563 LysAlaGly-----AsnGlnGlyMet 569
 761 CTGGCGCAGCAGTGGT---CCGGGTCCCGGA-----AGAGGTGACAGA----- 723
 570 ProGlyAspGlnGlyLeuProGlyProAlaGlyValLysGlyGlyGlyGlyGlyGlyGly 589
 722 ---GCAGGCGCTCCAGTGGAGTGAAGCAGCACCTGCGCCAGAGTCCAGCAGCCCGCC 666

```

Db 590 ProAlaGlySerGlnGlyAlaIleGlyAlaArgGlyProAlaGly---Thr 608
QY 665 CCAGGATGAGCAGTCCAGGCTCCGGGATCCGGGACAGCAGCCCTGTAGCC 606
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 609 ProGly-----ProAspGlyGlyGlyGlyGlyGlyGlyGlyValGly 625
QY 605 AGCCGG-----CCCTTGGGA-----TGAGAAAGAGGCTCAGCAGTCCCA 564
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 AlaAlaGlyHisGlnGlyProGlyGlyMetProGlyGlyGlyGlyGlyThrPro 645
QY 563 AGCAGAGTCCAGATGAAGCGCGCGCGCCATAGGTCCTCCAGCCAGTGGTCACTGG 504
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 Gly-----ProGlyGlyGlyGlyGlyGlyGlyHisArgGlyLeuGluGlyAsnMet 663
QY 503 CTGAGCTAGGAGCGGACACAGCAGCGCCAGCAGTGGACCAATGCCAGCACCATTGG 444
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 664 GlyArgAspGlyAlaArgGlyGlyProGlyPro----- 674
QY 443 TCATGAACCTCTCTCTACCCCACTT----- 417
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 -----SerGlyProProGlyProSerGlyAlaAsnGlyGluGlyGluSerGlySer 692
QY 416 -----CCAGCAGCAGCGCCACATAGTGTATCTCCGCGCCAAACACACCTCCAGCC 363
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 PheGlyProAlaGlyProAlaGlyLeuArg-----GlyProSerGly 706
QY 362 CAAAGTTAGCAGGTTGACCAAGAGCTGGGCTTCTCGGTCGCCAGCAGCGCGCTCA 303
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 GluArgGlyGlyGlyPro-----AlaGlyLeuProGlyPheAla-----GlyPro 722
QY 302 CCCACAGCTCTGGACCATAGTGGCGGCGGAGGCTCAGGCGCGCTCAGGAC 243
    ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 723 ProGlySerAspGly-GlnSerGlyProArgGly---GluGlyGlyProAlaGlyGly 741
QY 242 TCCAGAACCTCTCGTCTCGCTGCTGCTCCAGAGCTCGGCGCTCTCTCTCTGCTGCG 183
    ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 741 sGlyAspValGlyProAlaGlyProAlaGlyProSerGlyGlnSerGlyProSerGlyAl 761
QY 182 CCAACTGCTTAGGAATACAGCAGCGCGCCATTTCTGCCAGCCCTTGTGGCGGTCACGC 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 761 aSerGlyProAlaGlyProProGly-----GlyArgGlyAspAl 774
QY 122 TTCTACGCCCATGCTCAACA-----CTGTGCTGTGGGGCACCTCAGTGGGACACGT 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 774 aGlyProSerGlyLeu-ThrGlyPheProGlyAlaAlaGlyArgValGlyGly----- 791
QY 68 CTCATCACTCAGATCTGCGCGCGCGCGCTGTCACCGCGGACCGCGCG 16
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 792 -----ProGlyProAlaGlyIleAlaGlyProProGlySerAla 804
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 32
CA2B_MOUSE
ID CA2B_MOUSE STANDARD; PRT; 1650 AA.
AC 064739; 0921W0.
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 2(XI) chain precursor.
GN COL11A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,
RA Hall J., Lasky S., Hood L.;
RT *Sequence of the mouse major histocompatibility locus class II
RT region.*;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-1592 FROM N.A.

```

```

RC STRAIN=FVB/N, and 129/Sv; TISSUE=Cartilage;
RX MEDLINE=97135795; PubMed=8981332;
RA Vandenberg P., Vuorio M.M., Ala-Kokko L., Prockop D.J.;
RT "The mouse colla2 gene. Some transcripts from the adjacent rxr-beta
RL gene extend into the colla2 gene.";
CC Matrix Biol. 15:359-367(1996).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS (BY SIMILARITY).
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(XI) CAN ALSO BE FOUND INSTEAD
CC OF ALPHA 3(XI)=1(XI) (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -!- PM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF100956; AAC69905.1; -
DR EMBL; U16789; AAC67751.1; -
DR EMBL; U16790; AAC67752.1; -
DR MGD; MGI:88447; Coll1a2.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; Lamg; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Alternative splicing; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1414
FT PROPEP 1415 1650
FT DOMAIN 23 255
FT DOMAIN 487 1414
FT DOMAIN 1415 1650
FT CONFLICT 450 450
FT CONFLICT 618 619
FT CONFLICT 711 711
FT CONFLICT 757 759
FT CONFLICT 768 768
FT CONFLICT 790 790
FT CONFLICT 803 803
FT CONFLICT 836 836
FT CONFLICT 919 919
FT CONFLICT 1167 1167
FT CONFLICT 1300 1300
FT CONFLICT 1436 1436
SQ SEQUENCE 1650 AA; 162072 MW; 8ECDB8702E71E2DA CRC64;
    POTENTIAL
    COLLAGEN ALPHA 2(XI) CHAIN.
    CARBOXYL-TERMINAL PROPEPTIDE.
    NONHELICAL REGION.
    TRIPLE-HELICAL REGION.
    NONHELICAL REGION.
    R -> L (IN REF. 2).
    NQ -> KP (IN REF. 2).
    V -> A (IN REF. 2).
    TGP -> HGS (IN REF. 2).
    A -> S (IN REF. 2).
    R -> G (IN REF. 2).
    G -> V (IN REF. 2).
    G -> D (IN REF. 2).
    E -> V (IN REF. 2).
    P -> S (IN REF. 2).
    A -> T (IN REF. 2).
    I -> M (IN REF. 2).

```

```

Alignment Scores:
Pred. No.: 3.59e-07
Score: 307.00
Length: 1650
Percent Similarity: 31.06%
Conservative: 315
Best Local Similarity: 26.16%
Mismatch: 379
Query Match: 4.93%
Indels: 454
Gaps: 73
DB: 1

```

US-09-759-143-110 (1-3410) x CA2B_MOUSE (1-1650)

QY	3053	CCAAACAGGCTGTGGAGCTGGTGGCGCAAGATTGGGGTAGAGGAAAGTTGGGG---GTA	2997
Db	537	ProGlyValArgGlyMetAspGlyProHisGlyProLysGlySerLeuGlyProGln	556
QY	2996	GGGAAATTTGGCAGCTGCCTTCATCAGCCCATCTACAGAGAGTAGAGGAGTGGGA	2937
Db	557	GlyGlu-----ProGlyProGlyGlnGlnGlyThrPro	568
QY	2936	AGTGGGGGAACACGCGTGGCGCAAGA-----GAAGAGGGGTGG	2898
Db	569	GlyAlaGlnGlyLeuProGlyProGlnGlyAlaIleGlyProHisGlyGluLysGlyAla	588
QY	2897	TTAGGGAAGCGTTGAGACCTGAAGCCCCACCCTCTAC----CTTCCTTCAACACCCATA	2842
Db	589	ArgGlyLysProGlyLeuProGlyMetProGlySerAspGlyLeuProGlyHisPro---	607
QY	2841	CCTTGGGCTAACAGCATTTGGAATTATCATTTGGGATGAGTAGAATTTCCAAGCTCTCGGG	2782
Db	608	-----GlyLysGluGlyProPro	613
QY	2781	TTAGGCATTTTGGGGGCCAGACCCAGGAGAAAGATTTCTGCAATGATCAGCCCAAT	2722
Db	614	---GlyThrLysGlyAsnGlnGlyPro-----SerGlyProGlnGlyPro---	627
QY	2721	GACCAGCTATCTCAGGGGACCTGATGTTGGGATCCCCACCCTACCCAAATATT---	2666
Db	628	-----LeuGlyThrProGlyProArgGlyValLysGly	638
QY	2665	-----AGACACCAACACAGAAAGCTAGCAATGATTCCT	2630
Db	639	ValAspGlyIleArgGlyLeuLysGlyHisLysGlyGluLysGlyGluAspGlyPhePro	658
QY	2629	TCTACTTTGTAATAATAAAGTTAAATATTTAAATGCCCTGTCTGTGTGATGGCAACA	2570
Db	659	Gly-----PheLysGlyAspIleGlyValLysGlyAsp	669
QY	2569	GAAGGACCAACAGCCACATCCTGATAAAGGTAAGAGGGGTGATCAGCAAAAAGAC	2510
Db	670	ArgGlyGluValGlyValProGlySerArgGlyGluAspGlyProGlyLysProLysGly	689
QY	2509	AGTGCTGTGGCTGAGGGGACCTGGTCTTGTGTGTGTGCCCHCAGGACTCTTCCCTAC	2450
Db	690	ArgThrGlyProThrGlyAsp-----ProGlyProThrGlyLeuMetGlyGlu	705
QY	2449	AAATAAGTCATATGTTCAATCCATCCATGGAGAGTGTTTCATCTCAGAACTCCCATGCAA	2390
Db	706	Lys-----LysSerGly	706
QY	2389	GACCTACATAACGAGCTGCAGGTTAAGGGCTTAGAGATGGGAACACGAGTGACTGA	2330
Db	707	-----GlyLysLeuGlyValProGlyLeu---ProGlyThrProGly---	719
QY	2329	GTTTATTTCAGCTCCCAAAACCCCTTCTAGGTGTCTCAACTAGGAGCTAGCTGTTA	2270
Db	720	---ArgGlnGlyProLysGly-----	725
QY	2269	ACCTGAGCTGGGTAAATCCA-----CCTGCAGAGTCCCGCATTCAGTGCATGGAGC	2216
Db	726	-----SerLeuGlyPheProGlyPheProGlyAlaSerGlyGluLysGlyAlaArgGly	743
QY	2215	CCTTCTGGCTCCCTGTATAAGTCCAGACTGAACCCCTTGGAGGCCTCCAGTCAGGC	2156
Db	744	LeuSerGly-----LysSerGly	749
QY	2155	AGCCCTAGAGACTGGGGAGAGAGGAGGACGCCCGCCAGCCCGCTGTGCACTACGC	2096
Db	750	---ProArg-----GlyGluArg-----GlyProThrGlyProArgGlyGlnArgGly	764
QY	2095	ACCTCAGCAGCAGGGTGGCAGCAGAGCCACATTTCTGGCAGCAACAGAACTGG	2036
Db	765	ProArgGlyAlaThrGlyLysSerGlyAlaLysGlyThrSerGlyLys---	781

QY	2035	CGGCCAGCCCGCAGCCCGCATGGGGCTTAACAGGACGGGAGCTGGGA-----CCCACTGAG	1979
Db	782	-----GlyProHisGlyPro-----ProGlyGluArgGlyLeuProGlyPro	795
QY	1978	GCAGGCCCT-----CCACCCCAANTGCTGCTGAAGCTTTTCTA	1943
Db	796	GlnGlyProAsnGlyPheProGlyProGlyProGlyProGlyProAlaGlyLys-----	813
QY	1942	CGCTGAGTATTGGCCCAAGTCGCTCTGTCAATACTACCTGTGTAGCAAGATAAATGGC	1883
Db	814	-----AspGly-----AspGly	815
QY	1882	GA----CCAGNACCAGCGCTCGGCCACACACCATTATAGGCAGTCAGACTGCGCTGAGCT	1827
Db	816	LeuProGlyHisProGlyGlnArgGlyGluValGlyPheGlnGlyThrGlyProPro	835
QY	1826	GCACAATGGAGCCCAATAACAGGATGGGCCACCTGGGACAGCAGGAAGGCACCTATCCA	1767
Db	836	Gly-----ProProGlyValValGlyProGlnGlyThrAlaGlyGluSerGlyPro	852
QY	1766	---GGATGGCGAGTCCAGGCAGATGCCCGGCCCGGAACACCCTGGCTCGGTGGGCT	1710
Db	853	MetGlyGluArgGlyHisSerGlyProProGly-----Pro-ProGlyGluGlnGlyLe	870
QY	1709	CACCCACCACACAGTACGAGACATCACAGGCAGAG-----GCCCGCG	1665
Db	870	u-ProGlyThrSerGlyGlyGlnGlyThrGlyLysAspProGlyProProGlyAlaProG	890
QY	1664	AGCGCGGGTGGAGGTGGGAGCAGGCCACTGGCTCCAGCACCCACGTCGTCATTAGGA	1605
Db	890	lyLysAspGlyProAlaGlyLeuArgGlyPheProGlyGluArgGlyLeuProGlyThra	910
QY	1604	AGGAGCTCCAGGCTTA-----GGCGCTGGCAGGA	1575
Db	910	laGlyGlyProGlyLeuLyGlyAsnGluGlyProAlaGlyProProGlyPro-AlaGly	929
QY	1574	AGC-----TGTCATCAGGCTGCTCTCACTGCTAGCACCTCCAGTGT	1533
Db	930	SerProGlyGluArgGlyAlaAlaGlySerGlyPro-----	942
QY	1532	CCCTCGGTATTGGGCAGGAACACCTGCTTCCCGGTGTAAGGGAGGCCAGTGTGT	1473
Db	943	-----IleGlyProProGlyArgProGlyProGlnGlyProPro-----	955
QY	1472	AGGCAGGATCTGACGGCTGAGAAGGTGAACCGGTAGCGGGCTGAAGCTGTCAACA	1413
Db	956	-----GlyAlaAlaGlyGluLys-GlyValProGlyGluLysGly--ProIleGlyPro	972
QY	1412	CGGCCACACTGTGGACAGGCATGTGCACACCGCACGCCAGGGAAGCTGCCACACTGG	1353
Db	973	ThrGlyArgAspGlyVal-----GlnGlyProValGly--	983
QY	1352	CCAATAGACTGCTCGAGTCCGGAATCGTGCACACCGCGGTCCATGACCAGAGAGA	1293
Db	984	-----LeuProGlyProAlaGlyPro-----	990
QY	1292	CCAGGAGATGGCGCACTGCAGGAACA-----GCCCGCAGC	1257
Db	991	ProGly-----ValAlaGlyGluAspGlyLysGlyGluValGlyAspProGly	1007
QY	1256	TGCCCATCCGACGCCTTCATATGATGTCTCCGGGCTCGGTGCCGGCTCAGCTCTGG	1197
Db	1008	GlnLysGlyThrLysGlyAsnLysGlyGluHisGlyPro-----	1020
QY	1196	GCAGCCCTCGTACAGGCCCTCGCCACAGAAATCCGTGTAAACACGCTGAAGGTCATGA	1137
Db	1021	-----ProGly--ProProGlyProIleGlyPro-----	1029
QY	1136	GTGCCATCCAGCTGCACAGCTCAGCCACGAAGAGCGGCGCAGGG-----TCGGGGGCA	1083
Db	1030	-----ValGlyGlnProGlyAlaAlaGlyAlaAspGlyGluProGlyAla	1044

RT "Identification of a distinct type IV collagen alpha chain with
RT restricted kidney distribution and assignment of its gene to the
RT locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
RN [6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D.,
RA Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
RT to the region of the X chromosome containing the Alport syndrome
RT locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
RN [7]
RP SEQUENCE OF 1-20 FROM N.A.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,
RA Marynen P.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 1258-1270 FROM N.A. (SPliced FORM).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
RA Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood
RT cells: a complex mutation in the COL4A5 gene of an Alport patient
RT deletes the NCI domain.";
RL Kidney Int. 44:1316-1321(1993).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
RN [10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,
RA Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a
RT conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18(1991).
RN [11]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,
RA Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5
RT (IV) chain associated with X-linked Alport syndrome: characterization
RT of the mutation by direct sequencing of PCR-amplified lymphoblast
RT CDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
RN [12]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,
RA Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to
RT glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
RN [13]
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
RA Tryggvason K., Haggama-Schouten W.A.G., Roodvoets A.P., Rascher W.,
RA van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of
RT patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
RN [14]
RP VARIANTS AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872
RP AND C-1241.
RX MEDLINE=95322976; PubMed=7599631;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;

RT "Detection of 12 novel mutations in the collagenous domain of the
RT COL4A5 gene in Alport syndrome patients.";
RL Hum. Mutat. 5:197-204(1995).
RN [15]
RP VARIANT AS ARG-1649.
RX MEDLINE=96213750; PubMed=8651292;
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
RA Denison J.C., Fain P.R., Gregory M.C.;
RT "A mutation causing Alport syndrome with tardive hearing loss is
RT common in the western United States.";
RL Am. J. Hum. Genet. 58:1157-1165(1996).
RN [16]
RP VARIANTS AS.
RX MEDLINE=96213754; PubMed=8651296;
RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
RA Savi M., Ballabio A., de Marchi M.;
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
RT exons of the COL4A5 gene.";
RL Am. J. Hum. Genet. 58:1192-1204(1996).
RN [17]
RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
RP MET-1428.
RX MEDLINE=97094179; PubMed=8940267;
RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jäcasier D.,
RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
RA Gubler M.-C., Antignac C.;
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
RT syndrome.";
RL Am. J. Hum. Genet. 59:1221-1232(1996).
RN [18]
RP VARIANT AS ASP-1498.
RX MEDLINE=96233932; PubMed=8829632;
RA Tverskaya S., Bobryna V., Tsalykova F., Ignatova M.,
RA Krasnopol'skaya X., Evgrafov O.;
RT "Substitution of Al498D in noncollagen domain of alpha 5(IV) collagen
RT chain associated with adult-onset X-linked Alport syndrome.";
RL Hum. Mutat. 7:149-150(1996).
RN [19]
RP VARIANT AS GLN-1677.
RX MEDLINE=97295089; PubMed=9150741;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT "Common ancestry of three Ashkenazi-American families with Alport
RT syndrome and COL4A5 R1677Q.";
RL Hum. Genet. 99:681-684(1997).
RN [20]
RP VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.
RX MEDLINE=98112435; PubMed=9452056;
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
RA Trivelli A., Pinciaroli A.R., Ragalolo M., Rizzoni G.F., de Marchi M.;
RT "Missense mutations in the COL4A5 gene in patients with X-linked
RT Alport syndrome.";
RL Hum. Mutat. Suppl. 1:S106-S109(1998).
RN [21]
RP VARIANTS AS V-420; 456-P-P-458 DEL; D-573; D-624; D-635; 802-G--P-807
RP DEL; R-869; C-941; S-1030; S-1066; D-1143; R-1196; E-1261; S-1357
RP AND R-1649.
RX MEDLINE=99063529; PubMed=9848783;
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
RA Barker D.F., Gregory M.C., Atkin C.L., Strykarsdottir U., Neumann H.,
RA Springace J., Shows T.B., Pettersson E., Tryggvason K.;
RT "High mutation detection rate in the COL4A5 collagen gene in suspected
RT Alport syndrome using PCR and direct DNA sequencing.";
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
RN [22]
RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1158;
RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
RX MEDLINE=20030197; PubMed=10561141;
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
RT "Detection of mutations in the COL4A5 gene in over 90% of male

RT patients with X-linked Alport's syndrome by RT-PCR and direct
 RL sequencing.";
 RN Am. J. Kidney Dis. 34:854-862(1999).
 RP [23]
 RX VARIANT AS ARG-822.
 MEDLINE=20025011; PubMed=10563487;

Alignment Scores:
 Pred. No.: 3,59e-07 Length: 1685
 Score: 307.00 Matches: 286
 Percent Similarity: 30.64% Conservative: 63
 Best Local Similarity: 25.11% Mismatches: 412
 Query Match: 4.93% Indels: 379
 DB: 1 Gaps: 63

US-09-759-143-110 (1-3410) x CA54_HUMAN (1-1685)

QY 3039 GGAGCTGTGGGGAAGTTGGGGTAGGGGAAGTTGGGGTAGGGGAATTTTGGGCAG 2980
 DB 283 GlyGluLysGlyGluLysGlyGluGlnGlyGluProGlyLysArgGlyLysProGlyLys 302
 QY 2979 TGCCTTCATCAGCCAGTCCTAGAGAGAGTAGAGGGAGTGAAGTGGGGGAACACAGGC 2920
 DB 303 AspGlyGluAsnGlyGlnPro-----GlyIleProGly 313
 QY 2919 TGGGCCAAGAGAAGAGGGGTGGTTAGGAAGCCGTTGAGA----- 2880
 DB 314 Leu-ProGlyAspProGlyTyProGlyGluProGlyArgAspGlyGluLysGlyGlnLys 333
 QY 2879 -----CCTGAAGCCCAACCCCTTACCTTCCACACCCCTAACCTTGGGTA 2833
 DB 333 sGlyAspThrGlyProGlyProGlyProGlyLeuValIleProArgPro-----Gly-- 350
 QY 2832 ACAGCATTTGGAATATCATTTGGGATCAGTAGAATTTCCAGGTCCTGGTGGTAGGCATT 2773
 DB 351 Thr-----GlyIleThrIleGlyGluLysGlyAsnIleGlyLeuProGlyLeu----- 366
 QY 2772 TTGGGGGCCAGAGCCAGGAGAGAAGATTCCT-----GGCAATGATCAGCCC 2725
 DB 367 -----ProGlyGluLysGlyGluArgGlyPheProGlyIleGlnGlyPro 381
 QY 2724 AATGA-----CCAGTATCTCAGGGACCTGATTTGGGATCCCTACCCCAAT 2669
 DB 382 ProGlyLeuProGlyProGlyAlaAlaValMetGlyProGlyPro----- 398
 QY 2668 ATTAGACACCAACACAGAAAAGCTAGCAATGATTCCTTCTACTTTTAAATAATAA 2609
 DB 399 -----ProGlyPheProGlyGluArgGlyGlnLys--- 408
 QY 2608 GTTAATATTAATGCTGTCTCTGTGATGGCAACAGAGGACCAACAGCCACATC 2549
 DB 409 -----GlyAspGluGlyProProGly---Ile 416
 QY 2548 CTGATAAAGTAAGAGGGGTGGATCAGCAAAAGACAGTGTGGGCTCAGGGGAC 2489
 DB 417 SerIleProGlyProGlyProGlyLeuAspGlyGlnPro-----GlyAlaProGly 432
 QY 2488 CTGGTTCTGTGTGCTTCCCTCAGGA-----CTCTTCCCTACAAATAAGTCATATGT--- 2435
 DB 433 LeuProGlyProGlyProGlyProGlyProHisIleProSerAspGluIleCysGlu 452
 QY 2434 -----TCAATTCCTAGGAGGAGTGTTCATCTCTAGAAACTCCC 2396
 DB 453 ProGlyProGlyProGlyProGlySerProGlyAsp----- 464
 QY 2395 ATGCAAGAGCTACATTAACAGAGCTCCAGGTTAAGGGCTTAGA----- 2351
 DB 465 -----LysGlyLeuGlnGlyGluGlnGlyVal 473
 QY 2350 GATGGGAACACAGTGACTGAGTTATTACAGT-----CCC 2315
 DB 474 LysGlyAspLysGlyAspThrCysPheAsnCysIleGlyThrGlyIleSerGlyProPro 493

QY 2314 AAAAACCCCTTCTCTA---GGTGTGTCTCAACTAGGAGCTAGCTGTAAACCCCTGAGCCTG 2258
 DB 494 GlyGlnProGlyLeuProGlyLeuProGlyProGly-----SerLeu 508
 QY 2257 GGTAAATCCA-----CCTGCAGAGTCCCCGCATTCAGTCCATGGAGCCCTTCTGGCCTC 2204
 DB 509 GlyPheProGlyGlnLysGlyGluLysGlyGlnAlaGlyAlaThrGlyProLysGlyLeu 528
 QY 2203 CCTGTATTAAGTCCAGACTGAACCCCTTGGAAAGGCTCCAGTCA----- 2159
 DB 529 Pro-----GlyIleProGlyAlaProGlyAlaPro 538
 QY 2158 GGCAGCCCTAGACTGGGAGAG----- 2135
 DB 539 GlyPheProGlySerLysGlyGluProGlyAspIleLeuThrPheProGlyMetLysGly 558
 QY 2134 -----AGGAGAGAGCCGCCAGCCAGCCAGCTGTGCAG---CTACGCCACTCAGCA 2087
 DB 559 AsplyGlyGlnLeuGlySerProGlyAlaProGlyLeuProGlyLeuProGlyThrPro 578
 QY 2086 GCACAGGGTGGCAGCAGAGAGCCACATTTACTTTGGCAGCAACAGAACTGGCGCCAGCC 2027
 DB 579 GlyGlnAspGlyLeu-----ProGlyLeuProGly-----ProLys 590
 QY 2026 CGCAGCCCATGGGCTAACAGAGCGGGGAGCTGGACCCAGTGAAGGAGCCCTCCA 1967
 DB 591 GlyGluProGlyGlyIleThrPheLysGlyGluArgGlyProPro--GlyAsnProGlyL 610
 QY 1966 CCCCAATGTCTGGAAGTTTCTACGCTGAGTATTGGCCAGTCGCTCTGTCAATATAC 1907
 DB 610 eu----- 610
 QY 1906 TACCTGTGTAGCAAAAGTAATGCGCAGACCCAGCCCTGCGGACAGACACCATATAGGC 1847
 DB 611 -----ProGlyLeu----- 613
 QY 1846 AGTGACAGACTGCTGAGCTGGACATGAGGCCATTAACAGGGATGAGGCGCCACCTGGGA 1787
 DB 614 -----ProGlyAsnIleGlyProMetGlyPro--ProGlyPheGlyProProGlyP 630
 QY 1786 CAGCAGGAAGGACTATCCAGGATGCGAGGCTCCAGGAGATGCCCGG---CCGGAAC 1730
 DB 630 roValGlyGlu-----LysGlyIleGlnGlyValAlaGlyAsnProGlyGlnProGlyI 648
 QY 1729 CACCTTGGCTCGTGGGCTACCCACACACACACAGTACGAGACATACAGCAGAGGC 1670
 DB 648 leProGlyProLys--GlyAspProGlyGlnThrIleThrGlnProGlyLysProGlyLe 667
 QY 1669 CCGCAGAGCGGGTGGAGGTGGAGGAGCCAGCTGCTCCAGCAGCCACCTGTCATT 1610
 DB 667 uProGlyAsnProGlyArgAspGlyAspValGlyLeuProGlyAspProGlyLeuProG 687
 QY 1609 AGGAAGGAGCTCCAGGCTTAGGGCTTAGGAGGAGCTGTCATCAGCTGTCCTCATCT 1550
 DB 687 yGlnProGlyLeuProGlyIle--ProGlySerLysGlyGluProGlyIleProGlyI 706
 QY 1549 GCTAGCACCT---CCAGTGTCCCTCGGTATTTGGCAGGAACACCTGCTTCCCGGTG 1493
 DB 706 eGlyLeuProGlyProGlyProLysGly--PheProGlyIleProGlyProProGlyA 726
 QY 1492 GTAGGAGGAGGAGTGTGTAGGCGAGGATCTGCAGGCTGAGAGGTGTAACCCGCTGAG 1433
 DB 726 laProGlyThrPro-----G 731
 QY 1432 GCGGCTGAAGTGTCA-----CCAGGCCACACTGTGGGACAGGATGTGCCACCGC 1379
 DB 731 lyArgIleGlyLeuGluGlyProProGlyPro----- 741
 QY 1378 ACCCAGGGAAGTGCACACTGCCCAATAGACTGCTCAGTGCAGGCGCAATCGCTGCAC 1319
 DB 742 -----ProGlyPheProGlyProLysGlyGluProGlyPheAlaLeu 756
 QY 1318 CAGCGGCTCATGACAGAGAGAGACAGGAGGAGATGGCGCACTCCAGGAACA----- 1266

RA Apte S., Mattei M.-G., Olsen B.R.;
 RT "Cloning of human alpha 1(X) collagen DNA and localization of the
 RL COL10A1 gene to the q21-q22 region of human chromosome 6.";
 RN FBS Lett. 282:393-396(1991).
 [7]
 RP SEQUENCE OF 547-655 FROM N.A.
 RX MEDLINE-92077285; PubMed-1743401;
 RA Reichenberger E., Aigner T., von der Mark K., Stoeck H., Bertling W.;
 RT "In situ hybridization studies on the expression of type X collagen
 RL in fetal human cartilage.";
 RN Dev. Biol. 148:562-572(1991).
 [8]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97255959; PubMed-9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RL associated collagen (type IX), and network-forming collagen (type X)
 RN cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 Hum. Mutat. 9:300-315(1997).
 [9]
 RP VARIANTS SMCD ASP-598 AND PRO-614.
 RX MEDLINE-94136476; PubMed-8304336;
 RA Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,
 RT Grant M.E., Boot-Handford R.P.;
 RL "Amino acid substitutions of conserved residues in the
 RN carboxyl-terminal domain of the alpha 1(X) chain of type X collagen
 occur in two unrelated families with metaphyseal chondrodysplasia
 type Schmid.";
 Am. J. Hum. Genet. 54:169-178(1994).
 [10]
 RP VARIANTS SMCD ARG-591.
 RX MEDLINE-94272470; PubMed-8004099;
 RA McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;
 RT "Additional mutations of type X collagen confirm COL10A1 as the
 RL Schmid metaphyseal chondrodysplasia locus.";
 RN Hum. Mol. Genet. 3:303-307(1994).
 [11]
 RP VARIANTS SMCD VAL-618.
 RX MEDLINE-95181449; PubMed-7876225;
 RA Chan D., Cole W.G., Rogers J.G., Bateman J.F.;
 RT "Type X collagen multimer assembly in vitro is prevented by a Gly618
 RL to Val mutation in the alpha 1(X) NCI domain resulting in Schmid
 RN metaphyseal chondrodysplasia.";
 J. Biol. Chem. 270:4558-4562(1995).
 [12]
 RP VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.
 RX MEDLINE-95331767; PubMed-7607655;
 RA Bonaventure J., Chaminade F., Maroteaux P.;
 RT "Mutations in three subdomains of the carboxy-terminal region of
 RL collagen type X account for most of the Schmid metaphyseal
 RN dysplasias.";
 Hum. Genet. 96:58-64(1995).
 [13]
 RP VARIANTS SMCD PRO-600.
 RX MEDLINE-96375754; PubMed-8782043;
 RA Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,
 RT Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;
 RL "Mutations within the gene encoding the alpha 1(X) chain of type X
 RN collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but
 not several other forms of metaphyseal chondrodysplasia.";
 J. Med. Genet. 33:450-457(1996).
 [14]
 RP VARIANTS SMCD GLU-18 AND ARG-18.
 RX MEDLINE-97220591; PubMed-9067753;
 RA Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;
 RT "Mutations in the N-terminal globular domain of the type X collagen
 RL gene (COL10A1) in patients with Schmid metaphyseal
 RN chondrodysplasia.";
 Hum. Mutat. 9:131-135(1997).
 [15]
 RP VARIANTS SMD GLU-595.
 RX MEDLINE-99057503; PubMed-9837818;
 RA Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Ohashi H.,

RA Nakamura Y.;
 RT "Mutation of the type X collagen gene 'COL10A1' causes
 RL spondylometaphyseal dysplasia.";
 RN Am. J. Hum. Genet. 63:1659-1662(1998).
 [16]
 RP VARIANT SMCD CYS-597.
 RX MEDLINE-99069781; PubMed-9852679;
 RA Sawai H., Ida A., Nakata Y., Koyama K.;
 RT "Novel missense mutation resulting in the substitution of tyrosine by
 RL cysteine at codon 597 of the type X collagen gene associated with
 RN Schmid metaphyseal chondrodysplasia.";
 J. Hum. Genet. 43:259-261(1998).
 CC -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC
 CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
 CC MINERALIZATION ZONES OF HYALINE CARTILAGE.
 CC -!- SUBUNIT: HOMOTRIMER.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE
 CC METAPHYSEAL CHONDRODYSPLASIA (SMCD), A DOMINANTLY INHERITED
 CC DISORDER OF THE OSSEOUS SKELETON. THE CARDINAL FEATURES OF THE
 CC PHENOTYPE ARE MILD SHORT STATURE, COXA VARA AND A WADDLING GAIT.
 CC RADIOGRAPHY USUALLY SHOWS SCLEROSIS OF THE RIBS, FLARING OF THE
 CC METAPHYSES, AND A WIDE IRREGULAR GROWTH PLATE, ESPECIALLY OF THE
 CC KNEES.
 CC -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SPONDYLOMETAPHYSEAL
 CC DYSPLASIA (SMD). SMD COMPRISES A HETEROGENEOUS GROUP OF HERITABLE
 CC SKELETAL DYSPLASIAS CHARACTERIZED BY MODIFICATIONS OF THE
 CC VERTEBRAL BODIES OF THE SPINE AND METAPHYSES OF THE TUBULAR BONES.
 CC -!- SIMILARITY: STRONG. TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
 CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X60382; CAA42933.1; -
 CC EMBL; X65120; CAA46236.1; -
 CC EMBL; X98568; CAA67178.1; -
 CC EMBL; AL121963; CAB87590.1; -
 CC EMBL; X68531; RAC60615.1; -
 CC EMBL; X58879; CAA41686.1; -
 CC EMBL; M74050; AAA61221.1; -
 CC EMBL; X72579; CAA51170.1; -
 CC EMBL; X72580; CAA51170.1; JOINED.
 CC PIR; S15826; S15826.
 CC PIR; S30086; S30086.
 CC PIR; A43901; A43901.
 CC PIR; S18249; S18249.
 CC PIR; S21856; S21856.
 CC PIR; S26396; S26396.
 CC Genew; HGNC:2185; COL10A1.
 CC MIM; 120110; -
 CC MIM; 156500; -
 CC MIM; 184250; -
 CC InterPro; IPR001073; Clq.
 CC InterPro; IPR000087; Collagen.
 CC Pfam; PF00386; Clq; 1.
 CC Pfam; PF01391; Collagen; 8.
 CC PRINTS; PR00007; COMPLEMENTC1Q.
 CC SMART; SM00110; Clq; 1.
 CC PROSITE; PS01113; Clq; 1.
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Cartilage; Collagen; Signal; Disease mutation; Polymorphism.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 680 COLLAGEN ALPHA 1(X) CHAIN.
 FT DOMAIN 19 56 NONHELICAL REGION (NC2).
 FT DOMAIN 57 519 TRIPLE-HELICAL REGION.
 FT DOMAIN 520 680 NONHELICAL REGION (NC1).


```

Db      453 -ProIleGly---ProProGlyIleProGlyPhePro----- 463
Qy      540 CGGGGGGGCCATAGCTCCACCCAGTGGTCTAGCTAGGAGCGGGACACAG 481
Db      464 -----GlySerLysGlyAs 468
Qy      480 ACCAGGCCCCAGCACTGGACCAATGCCAGCAGCATGGTCATGAACCTTCCTCTACCC 421
Db      468 pProGly-----SerProGlyProPr 475
Qy      420 ACTTCCAGCAGCAGCGGCACATAGGTGATGCTCGCGCCAAACACACCTCCAGGCCA 361
Db      475 oGlyProAlaGly---IleAlaThrLysGlyLeuAsnGlyProThrGlyProGlyPr 494
Qy      360 AAGGTTAGCAGGTTGACCAAGAGAGCTGGCTTCCTCGGTCCGCGCAGGCGGCTCACC 301
Db      494 o-----ProGlyProArgGlyHisSer-----GlyGluPr 504
Qy      300 CACAGCCTCTGGACCATAGTGGCCAGCGGGTCTAGGGCTCAGGGGGCCGCTTCAGGCACCTC 241
Db      504 oGlyLeuProGlyPro-ProGlyProPro-----GlyProGlyGlnAlaValMetP 522
Qy      240 CAGAA---CTGCTTCCTCTCGCTCTGCTCCAGAGCTCGCGCCCTCTCTCTCTGCTGCC 184
Db      522 roGluGlyPheIleLysAlaGlyGlnArgProSerLeuSerGlyThrPro-LeuValSer 541
Qy      183 GCCAATGCTAGGAATCAGCCAGCGCCGCTTCTGCGCAGCCCTTGTGTCGGCTCCAG 124
Db      542 AlaAsn---GlnGlyValThrGlyMetProValSerAlaPheThrValIleLeuSerLys 560
Qy      123 CTCTCAGCCCATGCTCAACCTGCTGCTGTGGGGCACCTCAGTGGGGACACGCTCTCAT 64
Db      561 AlaTyProAlaIleGlyThrProIleProPheAspLysIleLeuTyAsnArgGlnGln 580
Qy      63 CACTCAGATCCTGGCGGAGGC 43
Db      581 HisTyAspProArgThrGly 587

RESULT 35
CA1A_HUMAN
ID CA1A_HUMAN STANDARD; PRT; 680 AA.
AC Q03692;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(X) chain precursor.
GN COL10A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=92109659; PubMed=1764025;
RA Thomas J.T., Cresswell C.J., Rash B., Nicolai H., Jones T.,
RA Solomon E., Grant M.E., Boot-Handford R.P.;
RA "The human collagen X gene. Complete primary translated sequence and
RA chromosomal localization.";
RT Biochem. J. 280:617-623(1991).
[2]
SEQUENCE FROM N.A.
RX MEDLINE=93012005; PubMed=1397333;
RA Reichenberger E., Beier F., Luvàlle P., Olsen B.R., von der Mark K.,
RA Bertling W.M.;
RT "Genomic organization and full-length cDNA sequence of human collagen
RT X.";
RL FEBS Lett. 311:305-310(1992).
[3]
SEQUENCE FROM N.A.
RA Beier F., Lamm M.B., von der Mark K.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.

```

```

RA Williams S.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
[5]
RX SEQUENCE OF 52-680 FROM N.A.
RX MEDLINE=92267014; PubMed=1587271;
RA Apte S.S., Seldin M.F., Hayashi M., Olsen B.R.;
RT "Cloning of the human and mouse type X collagen genes and mapping of
RT the mouse type X collagen gene to chromosome 10.";
RL Eur. J. Biochem. 206:217-224(1992).
[6]
RX SEQUENCE OF 561-666 FROM N.A.
RX MEDLINE=91243838; PubMed=2037056;
RA Apte S., Mattei M.-G., Olsen B.R.;
RT "Cloning of human alpha 1(X) collagen DNA and localization of the
RT COL10A1 gene to the q21-q22 region of human chromosome 6.";
RL FEBS Lett. 282:393-396(1991).
[7]
RX SEQUENCE OF 547-655 FROM N.A.
RX MEDLINE=92077285; PubMed=1743401;
RA Reichenberger E., Aigner T., von der Mark K., Stoeb H., Bertling W.;
RT "In situ hybridization studies on the expression of type X collagen
RT in fetal human cartilage.";
RL Dev. Biol. 148:562-572(1991).
[8]
RX REVIEW ON VARIANTS
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
[9]
RX VARIANTS SMCD ASP-598 AND PRO-614.
RX MEDLINE=94136476; PubMed=8304336;
RA Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,
RA Grant M.E., Boot-Handford R.P.;
RT "Amino acid substitutions of conserved residues in the
RT carboxyl-terminal domain of the alpha 1(X) chain of type X collagen
RT occur in two unrelated families with metaphyseal chondrodysplasia
RT type Schmid.";
RL Am. J. Hum. Genet. 54:169-178(1994).
[10]
RX VARIANTS SMCD ARG-591.
RX MEDLINE=94272470; PubMed=8004099;
RA McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;
RT "Additional mutations of type X collagen confirm COL10A1 as the
RT Schmid metaphyseal chondrodysplasia locus.";
RL Hum. Mol. Genet. 3:303-307(1994).
[11]
RX VARIANTS SMCD VAL-618.
RX MEDLINE=95181449; PubMed=7876225;
RA Chan D., Cole W.G., Rogers J.G., Bateman J.F.;
RT "Type X collagen multimer assembly in vitro is prevented by a Gly618
RT to Val mutation in the alpha 1(X) NC1 domain resulting in Schmid
RT metaphyseal chondrodysplasia.";
RL J. Biol. Chem. 270:4558-4562(1995).
[12]
RX VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.
RX MEDLINE=95331767; PubMed=7607655;
RA Bonaventure J., Chaminade F., Maroteaux P.;
RT "Mutations in three subdomains of the carboxy-terminal region of
RT collagen type X account for most of the Schmid metaphyseal
RT dysplasias.";
RL Hum. Genet. 96:58-64(1995).
[13]
RX VARIANTS SMCD PRO-600.
RX MEDLINE=96375754; PubMed=8782043;
RA Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,
RA Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;
RT "Mutations within the gene encoding the alpha 1(X) chain of type X
RT collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but
RT not several other forms of metaphyseal chondrodysplasia.";
RL J. Med. Genet. 33:450-457(1996).

```

[14]
 RP VARIANTS SMCD GLU-18 AND ARG-18.
 RA MEDLINE=97220591; PubMed=9067753;
 RX Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;
 RT "Mutations in the N-terminal globular domain of the type X collagen
 gene (COL10A1) in patients with Schmid metaphyseal
 chondrodysplasia";
 RL Hum. Mutat. 9:131-135(1997).
 [15]
 RP VARIANTS SMD GLU-595.
 RX MEDLINE=99057503; PubMed=9837818;
 RA Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Ohashi H.,
 RA Nakamura Y.;
 RT "Mutation of the type X collagen gene 'COL10A1' causes
 spondylometaphyseal dysplasia";
 RL Am. J. Hum. Genet. 63:1659-1662(1998).
 [16]
 RP VARIANT SMCD CYS-597.
 RX MEDLINE=99069781; PubMed=9852679;
 RA Sawai H., Ida A., Nakata Y., Koyama K.;
 RT "Novel missense mutation resulting in the substitution of tyrosine by
 cysteine at codon 597 of the type X collagen gene associated with
 Schmid metaphyseal chondrodysplasia";
 RL J. Hum. Genet. 43:259-261(1998).
 CC -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC
 CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
 MINERALIZATION ZONES OF HYALINE CARTILAGE.
 CC -!- SUBUNIT: HOMOTRIMER.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE
 METAPHYSEAL CHONDRODYSPLASIA (SMCD), A DOMINANTLY INHERITED
 DISORDER OF THE OSSEOUS SKELETON. THE CARDINAL FEATURES OF THE
 PHENOTYPE ARE MILD SHORT STATURE, COXA VARA AND A WADDLING GAIT.
 CC RADIOGRAPHY USUALLY SHOWS SCLEROSIS OF THE RIBS, FLARING OF THE
 METAPHYSES, AND A WIDE IRREGULAR GROWTH PLATE, ESPECIALLY OF THE
 KNEES.
 CC -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SPONDYLOMETAPHYSEAL
 DYSPLASIA (SMD). SMD COMPRISES A HETEROGENEOUS GROUP OF HERITABLE
 SKELETAL DYSPLASIAS CHARACTERIZED BY MODIFICATIONS OF THE
 VERTEBRAL BODIES OF THE SPINE AND METAPHYSES OF THE TUBULAR BONES.
 CC -!- SIMILARITY: STRONG. TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
 CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
 CC
 CC This swiss-prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X60382; CAA42933.1; -
 CC EMBL; X65120; CAA46236.1; -
 CC EMBL; X98568; CAA67178.1; -
 CC EMBL; AL121963; CAB87590.1; -
 CC EMBL; S68531; AAC60615.1; -
 CC EMBL; X58879; CAA41686.1; -
 CC EMBL; M74050; AAA61221.1; -
 CC EMBL; X72579; CAA51170.1; -
 CC EMBL; X72580; CAA51170.1; JOINED.
 CC PIR; S15826; S15826
 CC PIR; S30086; S30086.
 CC PIR; A43901; A43901.
 CC PIR; S18249; S18249.
 CC PIR; S21856; S21856.
 CC PIR; S26396; S26396.
 CC Genew; HGNC:2185; COL10A1.
 CC MIM; 120110; -
 CC MIM; 156500; -
 CC MIM; 184250; -
 CC InterPro; IPR001073; Clq.
 CC InterPro; IPR000087; Collagen.

DR Pfam; PF00386; Clq; 1.
 DR Pfam; PF01391; Collagen; 8.
 DR PRINTS; PRO0007; COMPLENTC1Q.
 DR SMART; SM00110; Clq; 1.
 DR PROSITE; PS01113; Clq; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Cartilage; Collagen; Signal; Disease mutation; Polymorphism.
 FT SIGNAL 1 18
 FT CHAIN 19 680
 FT DOMAIN 19 56
 FT DOMAIN 57 519
 FT DOMAIN 520 680
 FT DOMAIN 545 680
 FT VARIANT 18 18
 FT VARIANT 18 18
 FT VARIANT 545 545
 FT VARIANT 591 591
 FT VARIANT 591 591
 FT FTID=VAR_001841.
 Alignment Scores:
 Pred. No.: 5,09e-07 Length: 680
 Score: 304.50 Matches: 184
 Percent Similarity: 34.57% Conservative: 39
 Best Local Similarity: 28.53% Mismatches: 232
 Query Match: 4.74% Indels: 190
 DB: 1 Gaps: 38
 US-09-759-143-110 (1-3410) x CALA_HUMAN (1-680)
 QY 129 CCGGACCAAGGCTGGCAGAAATGGCCCTTCTAGCAGTTCCTAGCAGTTCGGCGCAG 188
 DB 74 ProGlyProSerGly---ProProGlyLysProGly-----TyrGlySerProGlyLeu 90
 QY 189 CAAGGAGGAGGCGCGCAGCTTCTGGAGCAGACGACGACGACGA---AGCAGTTCCTGGAGTG 245
 DB 91 GlnGlyGluProGlyLeuProGlyProProGlyProSerAlaValGlyLysProGlyVal 110
 QY 246 CCTGAAGCGCCCTGAGCCCTACCCGCTGCGCCACTATGTCCTGAGAGCT----- 297
 DB 111 ProGlyLeuProGlyLysProGlyGluArgGlyPro-TyrGlyProLysGlyAspValG1 130
 QY 298 -----GTGGTGAGCGCCCTCTCGCGCACCGCA 326
 DB 130 yProAlaGlyLeuProGlyProAArgGlyProProGlyProGlyProGlyProAl 150
 QY 327 AAGCCAGCTCTGCTGCTCAACCTGCTAACCTTTGGCTGGAGCTGTGTTGCCCGCAG 386
 DB 150 aGlyIleSerValProGlyLysProGlyGlnGlnGlyProThrGlyAlaProGlyProAr 170
 QY 387 GCATCACCTAATGTGCCGCTCTGCTGTGA-----AGTGGGGTAGAGGAGA 434
 DB 170 gGlyPheProGlyGluLysGlyAlaProGlyValProGlyMetAsnGlyGlnLysGlyG1 190
 QY 435 AGTTCATGACCATGGTGGGTCATGCTCCAGTCTGCTGGGCTGTGCTG-----TGTC 488
 DB 190 uMet-----GlyTyrGlyAlaProGlyArgProGlyGluArgGlyLeuPr 205
 QY 489 CGCTCTAGGCTCAGCCAGTACCAGTCCGCTGAGCAGCTATGGCGCGCGCGCCCTTCA 548
 DB 205 oglyProGlnGlyProThrGlyProSerGly-----ProProGlyValG1 220
 QY 549 TCTGGGCACTGTCCTTGGGCATCTGCTGAGCCT---CTTTCTCATCCCAAGGCCGCT 605
 DB 220 yLysArgGlyGluAsnGlyValProGlyGlnProGlyIleLysGlyAspArgGlyPhePr 240
 QY 606 GGCAGCAGGCTGCTGTG-----CCCGGATCCCGCCCTGGAGCTGGCAGCTGCTCA 659
 DB 240 oclyGluMetGlyProIleGlyProGlyProGlyProGlyPro----- 254
 QY 660 TCCTGGG---CGTGGGGCTGTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGG 716

```

Db 255 -ProGlyGluAArgGlyProGluGlyLeuGlyLysProGlyAlaAlaGlyAlaProGlyG1 274
QY 717 CCTGCTCTCTGACCT- - - - -CTTCCGGGA- - - - -CC 743
Db 274 nProGlyIle- - - - -ProGlyThrLysGlyLeuProGlyAlaProGlyIleAlaGlyPro 293
QY 744 CGGACCACTG- - - - -TGCCAGGCTACTCTGTCTATGCTTCCATGATCATGCTTG 794
Db 293 oGlyProProGlyPheGlyLysProGlyLeuProGlyLeuLysGlyGluAArgGlyProAl 313
QY 795 GGGGCTGCTGGCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 854
Db 313 aGlyLeuProGlyGlyProGlyAla- - - - -LysGlyGluGlnGlyProAlaGlyLe 330
QY 855 ACTTGGGACCCAGGA- - - - -GGAGTGCCTCTTTGGGCTGCTCACCTCATCTTCCTCA 908
Db 330 uProGlyLysProGlyLeuThrGlyProPro- - - - - 340
QY 909 CCGTGTAGCAGCACACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 968
Db 341 - - - - -GlyAsnMetGlyProGln- - - - - 346
QY 969 CAGAGGCTGTGCGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1028
Db 347 - - - - -GlyPro- - - - -LysGlyIleProGlySerHisG1 356
QY 1029 TTTTCCGGAAGCTGGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1088
Db 356 yLeuProGlyProLysGlyGluThrGlyProAlaGlyProAlaGlyTyPro- - - - - 373
QY 1089 GCACCTCGCGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1148
Db 374 - - - - -GlyAlaLysGlyGluAArgly- - - - - 380
QY 1149 CGCTGTTTACACGATTTCTGTCGGGAGGCTGTACACGGGCTGCTGCTGCTGCTGCTGCT 1208
Db 381 - - - - -SerProGlySerAspGlyLysProGlyTyProGlyLys- - - - -Pr 394
QY 1209 CGGGACCGAGCGCGGACACTATGATGAGCGCTTCGGATGGCGAGCTGGG- - - - - 1263
Db 394 oGlyLeuAspGlyPro- - - - -LysGlyAsnProGlyLeuProGlyProLy 409
QY 1264 -GCTGTCTCAGCTGGCCCATCTCCCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1322
Db 409 sGlyAspProGlyValGlyProGlyProGlyLeuProGlyProValGlyProAlaGlyAl 429
QY 1323 AGCGATTGCGGCACTCGAGCAGCTATTTGGCAGCTGTGGCAGCTTTCCCTGTGCTGCCG 1382
Db 429 a- - - - -LysGlyMetProGlyHisAsnGlyGlu- - - - -AlaGlyProAr 442
QY 1383 GTGCCACATGCTGCTCCACAGTGTGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1439
Db 442 gGlyAlaProGlyIleProGlyThr- - - - -ArgGly- - - - -ProIleGlyProProGlyIlePro 460
QY 1440 GTTTCACCTTCTCAGCCTCGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1499
Db 461 GlyPheProGly- - - - -SerLysGlyAspPro- - - - -GlySerProGlyProProG 476
QY 1500 AGAAGCAGGTGTTCTCTGCCCAATACCGAGGGACACTGGAGGTGCTAGCAGTGAGGACA 1559
Db 476 yProAlaGlyIleAlaThrLysGlyLeuAsnGly- - - - - 487
QY 1560 CCCTGTAGCAGCTTCTGCGAGGCTTAAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1619
Db 488 - - - - -ProThr-GlyPro- - - - -ProGlyProProGlyProArgGlyHis 501
QY 1620 TGGGTGTGGAGGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1679
Db 501 erGlyGluProGly- - - - -LeuProGlyProPro- - - - - 510
QY 1680 GTGATGCTCGTACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1739

```

```

Db 511 - - - - -GlyProProGlyPro- - - - -ProGlyG 518
QY 1740 GGGCATCTGCTGGACCTGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1799
Db 518 ln- - - - -AlaValMetProGluGlyPheIleLysAlaGlyGlnArgp 532
QY 1800 CATCCCTGTTTATGGGCTCCATGTCCTCCAGCTCAGCTCAGCTGCTGCTGCTGCTGCTG 1859
Db 532 roSerLeuSerGlyThrProLeuValSerAlaAsnGlnGlyValThrGlyMetProVal 552
QY 1860 CTGCCGAGCCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1913
Db 552 erAlaPheThrValIleLeuSerLysAlaTyProAlaIleGlyThrProIleProPhe 572
QY 1914 ACAAG 1918
Db 572 sPlys 573

```

RESULT 36

CA24_HUMAN

ID CA24_HUMAN STANDARD; PRT; 1712 AA.

AC P08572;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Collagen alpha 2(IV) chain precursor.

GN COL4A2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP MEDLINE=89086769; PubMed=3198637;

RA Hostikka S.L., Tryggvason K.;

RT "The complete primary structure of the alpha 2 chain of human type IV

RT collagen and comparison with the alpha 1(IV) chain.";

RL J. Biol. Chem. 263:19488-19493(1988).

RN [2]

RP SEQUENCE OF 1-1042 FROM N.A.

RC TISSUE=Placenta;

RX MEDLINE=88151998; PubMed=3345760;

RA Brazel D., Pollner R., Oberbauer I., Kuehn K.;

RT "Human basement membrane collagen (type IV). The amino acid sequence

RT of the alpha 2(IV) chain and its comparison with the alpha 1(IV)

RT chain reveals deletions in the alpha 1(IV) chain.";

RL Eur. J. Biochem. 172:35-42(1988).

RN [3]

RP SEQUENCE OF 1254-1712 FROM N.A.

RX MEDLINE=87219158; PubMed=3582677;

RA Hostikka S.L., Kurkinen M., Tryggvason K.;

RT "Nucleotide sequence coding for the human type IV collagen alpha 2

RT chain cDNA reveals extensive homology with the NC-1 domain of alpha 1

RT (IV) but not with the collagenous domain or 3'-untranslated region.";

RL FEBS Lett. 216:281-286(1987).

RN [4]

RP SEQUENCE OF 1451-1485 FROM N.A.

RX MEDLINE=87092438; PubMed=3025878;

RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;

RT "Human collagen genes encoding basement membrane alpha 1 (IV) and

RT alpha 2 (IV) chains map to the distal long arm of chromosome 13.";

RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).

RN [5]

RP SEQUENCE OF 1486-1712 FROM N.A.

RX MEDLINE=87250571; PubMed=2439508;

RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;

RT "Duplication of type IV collagen COOH-terminal repeats and species-

RT specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";

RL J. Biol. Chem. 262:9231-9238(1987).

RN [6]

RP SEQUENCE OF 1-33 FROM N.A.

RX MEDLINE=89034231; PubMed=3182844;

RA Soiminen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;

"The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region.";

J. Biol. Chem. 263:17217-17220(1988).

[7]

SEQUENCE OF 1-33 FROM N.A.

RP MEDLINE=89030632; PubMed=2846280;

RA Poeschl E., Pollner R., Kuehn K.;

"The genes for the alpha 1(IV) and alpha 2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure.";

EMBO J. 7:2687-2695(1988).

[8]

RP SEQUENCE OF 1-33 FROM N.A.

RC TISSUE=SKIN;

RX MEDLINE=93305049; PubMed=8317999;

RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;

"Identification of a novel sequence element in the common promoter region of human collagen type IV genes, involved in the regulation of divergent transcription.";

Biochem. J. 292:687-695(1993).

[9]

RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.

RC TISSUE=Placenta;

RX MEDLINE=89005112; PubMed=2844531;

RA Siebold B., Deutzmann R., Kuehn K.;

"The arrangement of intra- and intermolecular disulfide bonds in the carboxyterminal, non-collagenous aggregation and cross-linking domain of basement-membrane type IV collagen.";

Eur. J. Biochem. 176:617-624(1988).

CC

"- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/NIDOGEN.

CC

"- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) - ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.

CC

"- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.

CC

"- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC

"- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

EMBL; X05562; CAA29076.1;

DR EMBL; X05610; CAA29098.1;

DR EMBL; J02760; AAA58422.1;

DR EMBL; M36963; AAA53099.1;

DR EMBL; X12784; CAA31275.1;

DR EMBL; J04217; AAA53097.1;

DR PIR; A32024; A32024.

DR Genew; HGNC:2203; COL4A2.

DR MIM; 120090;

DR InterPro; IPR000087; Collagen.

DR InterPro; IPR001442; Procollagnc4.

DR Pfam; PF01391; Collagen; 24.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD000007; Collagen; 5.

DR ProDom; PD003923; Procollagnc4; 2.

DR SMART; SM00111; C4; 2.

DR

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KW Glycoprotein; Basement membrane; Collagen; Signal.

FT SIGNAL 1 25

FT PROPEP 26 183

FT CHAIN 184 1712

FT DOMAIN 184 1484

FT DISULFID 1485 1712

FT DISULFID 1504 1593

FT DISULFID 1537 1555

FT DISULFID 1549 1555

FT DISULFID 1612 1708

FT DISULFID 1646 1705

FT DISULFID 1658 1665

FT CARBOHYD 138 138

FT CONFLICT 471 471

FT CONFLICT 683 683

FT CONFLICT 1575 1575

FT CONFLICT 1663 1663

FT CONFLICT 1701 1701

FT CONFLICT 1712 1712

SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;

Alignment Scores:

Pred. No.:	4.86e-07	Length:	1712
Score:	304.00	Matches:	302
Percent Similarity:	28.95%	Conservative:	68
Best Local Similarity:	23.63%	Mismatches:	386
Query Match:	4.74%	Indels:	522
DB:	1	Gaps:	70

US-09-759-143-110 (1-3410) x CA2A_HUMAN (1-1712)

QY	95	GCAGCAGGTGTTGAGCATGGCTGAGACCTGGACCGGCACCAAGGGCTGCAGAAATG	154
Db	95	AlaProGlyValThr---GlyProLysGlyAspValGlyAlaArgValSerGlyPhe	113
QY	155	---GGCGCCTGGCTGATCTTAGCAGTGGCGGAGCAAGGA-----GGAGAGGCCGC	205
Db	114	ProGlyAlaAspGlyLeuPro-GlyHisProGlyGlnGlyGlyProArgGlyArgProG	133
QY	206	AGTCTTGGAGCAGACCGCAGAGCAGAGTCTGGAGTGGCTGACGACGCCCTC-----	260
Db	133	yTyrAspGlyCysAsnGlyThrGlnGlyAspSerGly---ProGlnGlyProGlySe	152
QY	261	-GAGCCCTAC-----CCGCTGGCCCACTATGTCTCAGAGCGCTGGGTGAGCGCCCTG	313
Db	152	rGlyPheThrGlyProProGlyPro-GlnGlyProLysGlyGlnLysGlyGluProT	172
QY	314	CTCGCGCACCG-----GAAAGCCCGAGCTCTTGTCTGGTCAACCTGTACCTTTGGCCCTG	367
Db	172	yrAlaLeuProLysGluGluArgAspArgTyrArgGlyGluProGlyGlu-----ProG	190
QY	368	GA-----GGTGTGTTTGGCGCAGGCATCACCTATGTGCGGCTCTGCTG	412
Db	190	lyLeuValGlyPheGlnGlyProGlyArgProGlyHisValGlyGlnMetGlyProV	210
QY	413	CTGGAAGTGGCGTAGA-----GGAGAGTTCATG	442
Db	210	alGlyAlaProGlyArgProGlyProGlyProGlyProLysGlyGlnGlnGlyA	230
QY	443	ACCATGTGCTGGGCATTTGG-----TCCAGTGTGGGC	475
Db	230	snArgGlyLeuGlyPheThrGlyValGlyGlyGlyGlyGlyGlyGlnProGlyP	250
QY	476	CTGGTCTGTCTCCGCTCTAGCTAGCTAGCCAG-----TGACCATCTGGCTGACGCTAT	529
Db	250	roAsnGlyIleProSerAspThrLeuHisProIleAlaProThrGlyValThrPheH	270
QY	530	GGCGCGCG-----	537
Db	270	isProAspGlnTyrLysGlyGlyGlyGlyGlyGlyGlyGlyGlyIleArgGlyIleS	290
QY	538	-----CGGCCCTTCAT-----CTGG	553

Qy	2225	ACTGGAATCGGGGACTCTGCGAGGTGCTATACCCAGGCTCAGGTTTACACAGCTAGCCTCC	2284
Db	908	yle-----AspGlyMetProGlyLeuLysGlyAspArgGlySerP	925
Qy	2285	TAGTTTGAGACACACCTAGAGAAGGTTTGGGAGCTGAATAAACTCAGCTCA--CCTGG	2341
Db	925	oGlyMetAsp-----GlyPheGlnGlyMetProGlyLeuLysGlyArgProGln	941
Qy	2342	TTTTCCCATCTCTAAGCCCTTAACTGCGAGCTCGHTTAAATGTAGCTCTTGCATGGGAGT	2401
Db	941	yPheProGly-----	944
Qy	2402	TTCTAGGATGAAACACTCTCTCCATGGCATTTGAACATATGACTTATTTCTAGGGGAAGAG	2461
Db	945	-----SerLysGlyGluAlaGlyPhePheGlyIleProGlyLeu-LysGlyLeuA	961
Qy	2462	TC-----CTGAGGGGCAACACACAGAACCCAGGTGCCCTCAGCCACACAGCACT	2509
Db	961	laGlyGluProGlyPheLysGlySerArgGlyAspProGlyPro-----	975
Qy	2510	GTCTTTTGTGTGATTCACACCCCTCTTACCTTTTATCAGGATGTGCGCTGTTGGTCTCTTC	2569
Db	976	-----ProGlyProProPro-----ValI	982
Qy	2570	TGTTGCCATCACAGACACAGGCAATTAATAATTTAACTATTATTAAACAACTAGA	2629
Db	982	leLeuProGlyMetLysAsp-----	988
Qy	2630	AGGGAATCCATCTGCTAGCTTTTCTGTGTTGGTGTCTAATATTGGGTAGGTTGGGGGATC	2689
Db	989	-----lleLysGlyGluLysGlyAsp-----	995
Qy	2690	CCCAACAATCAGTCCCTCGAGATAGCTGGTCAATGGGCTGATCATTGCCAGAACTTTCT	2749
Db	996	-----GluGlyProMet-----	999
Qy	2750	TCTCTGGGTCTGGCCCCCAAAATGCCTAACCCAGGACCTTGGAAATTTCTACTCATCC	2809
Db	1000	-----GlyLeu-----	1001
Qy	2810	CAAAATGAATTTCCAATGCTGTATTACCAAGTTAGGTGTTGAAGGAAGGTAGAGGGTG	2869
Db	1002	-----LysGlyTyrLeuG	1006
Qy	2870	GGGCTTCAGTCTCAACGGCTTCCCTAACCAACCCCTCTTCTCTTGGCCAGCCTGGTTC	2929
Db	1006	lyAlaLysGlyIleGlnGlyMetProGlyIlePro-----GlyLeuSerGlyIleP	1023
Qy	2930	CCCCATTCACCTCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACTGCCCAAA	2989
Db	1023	roGlyLeuPro-----GlyArgProGlyHisIleLysGlyValLysGlyA	1038
Qy	2990	TTTCCCCTACCCCAACTTTTCCCTACCCCAACTTTCCCAACCAAGCTCCACAAACCTGT	3049
Db	1038	spIleGlyValProGlyIleProGlyLeuProGlyPhePro-----	1051
Qy	3050	TTGGAGCTACTCGGACCAAGACCAAGTCGGGTTTCCCAAGCCTTTGTCCACTCTCA	3109
Db	1052	-----GlyValAlaGlyProProGlyIleThr--GlyPheProGlyPheIleGlySerAr	1069
Qy	3110	GCCCCCAGAGTATATCTGTCTGTTGGGAATCTCACACAGAAATCTCAGAGCACCCCTCG	3169
Db	1069	g-----GlyAspLysGlyAlaPro-----	1075
Qy	3170	CTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGG	3203
Db	1076	----GlyValAlaGlyLeuTyrGlyGluIleGly	1085

RESULT 37

CA1B HUMAN

CALB_HUMAN	
ID	CALB HUMAN
	STANDARD:
	PRT: 1806 AA.

ID	CA1B_HUMAN	STANDARD;	PRT;	180
AC	P12107	: Q14034:	Q9UIT4: Q9UIT5: Q9UIT6:	

AC P12107; Q14034; Q901T4; Q901T5
DT 01-OCT-1989 (Rel. 12, Created)

CC -!- DISEASE: Defects in COL1A1 are the cause of Stickler syndrome
 CC type 2 (STL2), also known as vitreous type 2, or beaded vitreous
 CC type, due to the presence of irregularly thickened fiber bundles
 CC throughout vitreous cavity. Stickler syndrome (STL) (hereditary
 CC progressive arthro-ophthalmopathy) is an autosomal dominant
 CC disorder characterized by progressive myopia beginning in the
 CC first decade of life, vitreo-retinal degeneration, retinal
 CC detachment, cleft palate, midfacial hypoplasia, osteoarthritis,
 CC and sensorineural hearing loss.
 CC -!- DISEASE: Defects in COL1A1 are the cause of Marshall syndrome, an
 CC autosomal dominant disorder with ocular, oro-facial, auditory and
 CC skeletal manifestations. It shares several features with Stickler
 CC syndrome, such as midfacial hypoplasia, high myopia, and
 CC sensorineural-hearing deficit.
 CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
 CC -!- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; J04177; AAA51891.1; -
 DR EMBL; AF101112; AAF04724.1; JOINED.
 DR EMBL; AF101079; AAF04724.1; JOINED.
 DR EMBL; AF101080; AAF04724.1; JOINED.
 DR EMBL; AF101081; AAF04724.1; JOINED.
 DR EMBL; AF101082; AAF04724.1; JOINED.
 DR EMBL; AF101083; AAF04724.1; JOINED.
 DR EMBL; AF101084; AAF04724.1; JOINED.
 DR EMBL; AF101085; AAF04724.1; JOINED.
 DR EMBL; AF101086; AAF04724.1; JOINED.
 DR EMBL; AF101087; AAF04724.1; JOINED.
 DR EMBL; AF101088; AAF04724.1; JOINED.
 DR EMBL; AF101089; AAF04724.1; JOINED.
 DR EMBL; AF101090; AAF04724.1; JOINED.
 DR EMBL; AF101091; AAF04724.1; JOINED.
 DR EMBL; AF101092; AAF04724.1; JOINED.
 DR EMBL; AF101093; AAF04724.1; JOINED.
 DR EMBL; AF101094; AAF04724.1; JOINED.
 DR EMBL; AF101095; AAF04724.1; JOINED.
 DR EMBL; AF101096; AAF04724.1; JOINED.
 DR EMBL; AF101097; AAF04724.1; JOINED.
 DR EMBL; AF101098; AAF04724.1; JOINED.
 DR EMBL; AF101099; AAF04724.1; JOINED.
 DR EMBL; AF101100; AAF04724.1; JOINED.
 DR EMBL; AF101101; AAF04724.1; JOINED.
 DR EMBL; AF101102; AAF04724.1; JOINED.
 DR EMBL; AF101103; AAF04724.1; JOINED.
 DR EMBL; AF101104; AAF04724.1; JOINED.
 DR EMBL; AF101105; AAF04724.1; JOINED.
 DR EMBL; AF101106; AAF04724.1; JOINED.
 DR EMBL; AF101107; AAF04725.1; JOINED.
 DR EMBL; AF101081; AAF04725.1; JOINED.
 DR EMBL; AF101082; AAF04725.1; JOINED.
 DR EMBL; AF101083; AAF04725.1; JOINED.
 DR EMBL; AF101084; AAF04725.1; JOINED.
 DR EMBL; AF101085; AAF04725.1; JOINED.
 DR EMBL; AF101086; AAF04725.1; JOINED.
 DR EMBL; AF101087; AAF04725.1; JOINED.
 DR EMBL; AF101088; AAF04725.1; JOINED.
 DR EMBL; AF101089; AAF04725.1; JOINED.
 DR EMBL; AF101090; AAF04725.1; JOINED.

DR EMBL; AF101091; AAF04725.1; JOINED.
 DR EMBL; AF101092; AAF04725.1; JOINED.
 DR EMBL; AF101093; AAF04725.1; JOINED.
 DR EMBL; AF101094; AAF04725.1; JOINED.
 DR EMBL; AF101095; AAF04725.1; JOINED.
 DR EMBL; AF101096; AAF04725.1; JOINED.
 DR EMBL; AF101097; AAF04725.1; JOINED.
 DR EMBL; AF101098; AAF04725.1; JOINED.
 DR EMBL; AF101099; AAF04725.1; JOINED.
 DR EMBL; AF101100; AAF04725.1; JOINED.
 DR EMBL; AF101101; AAF04725.1; JOINED.
 DR EMBL; AF101102; AAF04725.1; JOINED.
 DR EMBL; AF101103; AAF04725.1; JOINED.
 DR EMBL; AF101104; AAF04725.1; JOINED.
 DR EMBL; AF101105; AAF04725.1; JOINED.
 DR EMBL; AF101106; AAF04725.1; JOINED.
 DR EMBL; AF101107; AAF04725.1; JOINED.
 DR EMBL; AF101108; AAF04725.1; JOINED.
 DR EMBL; AF101109; AAF04725.1; JOINED.
 DR EMBL; AF101110; AAF04725.1; JOINED.
 DR EMBL; AF101111; AAF04725.1; JOINED.
 DR EMBL; AF101112; AAF04726.1; JOINED.
 DR EMBL; AF101079; AAF04726.1; JOINED.
 DR EMBL; AF101080; AAF04726.1; JOINED.
 DR EMBL; AF101081; AAF04726.1; JOINED.
 DR EMBL; AF101082; AAF04726.1; JOINED.
 DR EMBL; AF101083; AAF04726.1; JOINED.
 DR EMBL; AF101084; AAF04726.1; JOINED.
 DR EMBL; AF101085; AAF04726.1; JOINED.
 DR EMBL; AF101086; AAF04726.1; JOINED.
 DR EMBL; AF101087; AAF04726.1; JOINED.
 DR EMBL; AF101088; AAF04726.1; JOINED.
 DR EMBL; AF101089; AAF04726.1; JOINED.
 DR EMBL; AF101090; AAF04726.1; JOINED.
 DR EMBL; AF101091; AAF04726.1; JOINED.
 DR EMBL; AF101092; AAF04726.1; JOINED.
 DR EMBL; AF101093; AAF04726.1; JOINED.
 DR EMBL; AF101094; AAF04726.1; JOINED.
 DR EMBL; AF101095; AAF04726.1; JOINED.
 DR EMBL; AF101096; AAF04726.1; JOINED.
 DR EMBL; AF101097; AAF04726.1; JOINED.
 DR EMBL; AF101098; AAF04726.1; JOINED.
 DR EMBL; AF101099; AAF04726.1; JOINED.
 DR EMBL; AF101100; AAF04726.1; JOINED.
 DR EMBL; AF101101; AAF04726.1; JOINED.
 DR EMBL; AF101102; AAF04726.1; JOINED.
 DR EMBL; AF101103; AAF04726.1; JOINED.
 DR EMBL; AF101104; AAF04726.1; JOINED.
 DR EMBL; AF101105; AAF04726.1; JOINED.
 DR EMBL; AF101106; AAF04726.1; JOINED.
 DR EMBL; AF101107; AAF04726.1; JOINED.

Alignment Scores:

Pred. No.:	4.84e-07	Length:	1806
Score:	304.00	Matches:	313
Percent Similarity:	30.30%	Conservative:	56
Best Local Similarity:	25.70%	Mismatches:	410
Query Match:	4.88%	Indels:	441
DB:	1	Gaps:	71

US-09-759-143-110 (1-3410) x CA1B_HUMAN (1-1806)

QY	3101	ACAAAGGCTTGGGAACCGCACTTTGTCTTCTGCT	-----	3066
Db	409	ThrAspIleThrGluThrSerIleasnGlyHisGlyAlaTyrGlyGlyGlycInLys	428	
QY	3065	-----CCTGCAGTAGCT---CCAAACAGGGTTGTGGAGCTGGTGGGAAAAGTTGGGGGT	3015	
Db	429	GlyGluProAlaValValGluProGlyMetLeuValGlu	443	
QY	3014	AGGGGAAAGTTGGGGTAGGGGAAATTTTGGGCAGTGCCTTCATCAGCCAGTCCTAGAG	2955	
Db	444	ProGlyProAlaGlyProAlaGlyIleMetGly	459	

CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; 222964; CAA80536.1; -
 DR EMBL; 222964; CAA80537.1; -
 DR EMBL; J05066; AAR27989.1; -
 DR EMBL; J05066; AAR27989.1; -
 DR EMBL; U22327; AAR64312.1; ALT_SEQ.
 DR EMBL; U53342; AAA96215.1; -
 DR EMBL; U53342; AAA96216.1; -
 DR PIR; A34476; A34476.
 DR WormPep; F01G12.5a; CE04334.
 DR WormPep; F01G12.5b; CE04335.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR001442; ProcollagenC4.
 DR Pfam; PF01391; Collagen; 23.
 DR Pfam; PF01413; C4; 2.
 DR ProDom; PD000007; Collagen; 1.
 DR ProDom; PD003923; ProcollagenC4; 2.
 DR SMART; SM00111; C4; 2.
 DR Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
 KW Alternative splicing; Glycoprotein; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 1758
 FT COLLAGEN ALPHA 2(IV) CHAIN.
 FT 7S DOMAIN.
 FT 42 1527
 FT TRIPLE-HELICAL REGION.
 FT DOMAIN 1528 1758
 FT NONHELICAL REGION (NC1).
 FT DISULFID 1546 1635
 FT OR 1632 (BY SIMILARITY).
 FT DISULFID 1591 1597
 FT BY SIMILARITY.
 FT DISULFID 1654 1750
 FT OR 1747 (BY SIMILARITY).
 FT DISULFID 1688 1747
 FT BY SIMILARITY.
 FT DISULFID 1700 1707
 FT O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
 FT CARBOHYD 248 248
 FT VARSPLIC 229 264
 FT DIGANGPAGPPGPIASTMSKTIIGPKDLEKGEK (IN
 FT ISOFORM II).
 FT G -> E (IN MN114; 73% LETHAL).
 FT A -> T (IN MN126; 100% LETHAL).
 FT G -> E (IN MN109; 37% LETHAL).
 FT G -> R (IN MN103 AND MN151; 96% LETHAL).
 FT G -> R (IN MN152; 50% LETHAL).
 FT G -> R (IN MN101; 100% LETHAL).
 FT G -> E (IN MN129; 100% LETHAL).
 FT G -> E (IN MN143; 100% LETHAL).
 FT G -> R (IN G30; 90% LETHAL).
 FT G -> R (IN E1470; 94% LETHAL).
 FT G -> E (IN MN139; 20% LETHAL).
 FT G -> D (IN G25; 2% LETHAL).
 FT G -> D (IN MN147; 7% LETHAL).
 FT G -> D (IN G37 AND B246; 9% LETHAL).
 FT E -> D (IN REF. 3).
 FT CONFLICT 1604 1604
 FT CONFLICT 1682 1682
 FT P -> L (IN REF. 1 AND 3; AAA96216).
 FT SEQUENCE 1758 AA; 167750 MW; 97EE3F3DBB2D2AC5 CRC64;

Alignment Scores:
 Pred. No.: 5,1e-07
 Score: 303.50
 Percent Similarity: 31.55%
 Best Local Similarity: 24.96%
 Query Match: 4.73%
 DB: 1

Length: 1758
 Matches: 318
 Conservative: 84
 Mismatches: 435
 Indels: 438
 Gaps: 73

us-09-759-143-110 (1-3410) x CA24_CABEL (1-1758)


```

QY 2776 GCCTAACCCAGGACCTTGGAAATCTACTCATCCCAAAATGATTAATTCRAATGCTCTTAC 2835
Db 1332 ysArgGlyGluAsp--Gly-LeuProGlyValProGlyArgAspGlyGlnProGlyIle 1350
QY 2836 CCAAGGTAGGCTTCAAGGAAGTAGAGGGTGGGCTTCAGGTCTC----- 2883
Db 1351 ProGlyLeu-----LysGlyGluValGlyGlyAlaGlyLeuProGlyGlnPro 1366
QY 2884 -----AACGGCTTCCTAACCAACCCCTCTTCT 2910
Db 1367 GlyPheProGlyIleProGlyLeuLysGlyGluGlyGlyLeuProGlyPheProGlyAla 1386
QY 2911 CTTGGCCAGCCTGGTTCCTCCCACTTCCA----- 2940
Db 1387 LysGlyGluAlaGlyPheProGlyThrProGlyValProGlyTyrAlaGlyGluLysGly 1406
QY 2941 -----CPCCTCTACTCTCTAGGACTGGGCTGATGAAGGCACTGCCCAAAAT 2991
Db 1407 AspGlyGlyLeuProGlyLeuProGlyArgAspGlyLeu----- 1419
QY 2992 TCCCTACCCCACTTTCCTTCCCTACCCCACTTTCCTCCCACTCCCAACCCCTGTT 3051
Db 1420 ---ProGlyAlaAspGlyProValGlyProProGlyProSer----- 1432
QY 3052 GGAGCTACTGAGGACCAAGCAAGCAAGTGCCTTTCCTCCCACTCCCAACCCCTGAGC 3111
Db 1433 -----Gly 1433
QY 3112 CCCAGAGTATATCTGCTTGGGGAATCTCACACAGAACTCAGGAGCCTCCCTGCCT 3171
Db 1434 ProGlnAsnLeu-----ValGluProGlyGluLysGlyLeuPro 1446
QY 3172 GAGCTAAGGAGGCTTATCTCTCAGGGG 3201
Db 1447 GlyLeuProGlyAlaProGlyLeuArgGly 1456

RESULT 39
CA21.RAT
ID CA21.RAT STANDARD; PRT: 1372 AA.
AC F02456; Q9R1E8;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Guenther D., Seibold S., Marx M.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 86-98.
RC TISSUE=Skin;
RX MEDLINE=67162268; PubMed=5337886;
RA Kang A.H., Bornstein P., Piez K.A.;
RT "The amino acid sequence of peptides from the cross-linking region of
RT rat skin collagen";
RL Biochemistry 6:788-795(1967).
RN [3]
RP SEQUENCE OF 99-102.
RC TISSUE=Skin;
RX MEDLINE=69206881; PubMed=5785232;
RA Fietzek P.P., Piez K.A.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 2 chain of rat skin collagen.";
RL Biochemistry 8:2129-2133(1969).
RN [4]
RP SEQUENCE OF 102-144.
RC TISSUE=Skin;

```

```

RX MEDLINE=73049496; PubMed=4636752;
RA Fietzek P.P., Kell I., Kuehn K.;
RT "The covalent structure of collagen. Amino acid sequence of the N-
RL terminal region of alpha 2-CB4 from calf and rat skin collagen.";
RN FEBS Lett. 26:66-68(1972).
RN [5]
RP SEQUENCE OF 423-452.
RC TISSUE=Skin;
RX MEDLINE=71115216; PubMed=5544653;
RA Higberger J.H., Kang A.H., Gross J.;
RT "Comparative studies on the amino acid sequence of the alpha 2-CB2
RL peptides from chick and rat skin collagens.";
RN Biochemistry 10:610-616(1971).
RN [6]
RP SEQUENCE OF 453-501.
RC TISSUE=Skin;
RX MEDLINE=75059250; PubMed=4435743;
RA Fietzek P.P., Kuehn K.;
RT "The covalent structure of collagen: amino acid sequence of the N-
RL terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5
RL from calf skin collagen.";
RN Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).
RN [7]
RP SEQUENCE OF 791-836.
RC TISSUE=Skin;
RX MEDLINE=74055004; PubMed=4763308;
RA Fietzek P.P., Kuehn K.;
RT "The covalent structure of collagen: amino acid sequence of the N-
RL terminal region of alpha 2-CB5 from rat skin collagen.";
RN FEBS Lett. 36:289-291(1973).
RN [8]
RP ORDER OF CNBR PEPTIDES.
RX MEDLINE=70181852; PubMed=5443712;
RA Vuust J., Lane J.M., Fietzek P.P., Miller E.J., Piez K.A.;
RT "The order of the CNBR peptides from the alpha 2 chain of collagen.";
RL Biochem. Biophys. Res. Commun. 38:703-708(1970).
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF121217; AAD41775.1;
DR PIR; A02867; CGRT2S.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib-collagen_C.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Collagen; 3.
DR ProDom; PD002078; Fib-collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 24
FT PROPEP 25 85
FT FT Amino-terminal propeptide
FT (by similarity).
FT CHAIN 86 1108
FT PROPEP 1109 1372
FT FT Carboxyl-terminal propeptide
FT (by similarity).
FT SITE 783 785
FT SITE 828 830
FT SITE 1011 1013

```


FT MOD_RES 86 86 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
 FT MOD_RES 90 90 CONVERTED TO AN ALDEHYDE GROUP THAT IS
 FT CARBOHYD 1273 1273 INVOLVED IN CROSS-LINKING.
 FT CONFLICT 132 132 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 137 137 T -> P (IN REF. 4).
 FT CONFLICT 145 422 S -> P (IN REF. 4).
 FT CONFLICT 431 432 MISSING (IN REF. 4).
 FT CONFLICT 494 494 ST -> TS (IN REF. 5).
 FT CONFLICT 497 497 E -> Z (IN REF. 6).
 FT CONFLICT 502 790 N -> A (IN REF. 6).
 FT CONFLICT 825 825 MISSING (IN REF. 7).
 FT CONFLICT 825 825 R -> K (IN REF. 7).
 SQ SEQUENCE 1372 AA; 129564 MW; B069371A8DB20A72 CRC64;

Alignment Scores:
 Pred. No.: 6,1e-07 Length: 1372
 Score: 302.00 Matches: 302
 Percent Similarity: 31.04% Conservative: 55
 Best Local Similarity: 26.26% Mismatches: 393
 Query Match: 4.85% Indels: 402
 DB: 1 Gaps: 61

US-09-759-143-110 (1-3410) x CA21_RAT (1-1372)

QY 3074 GCTTCTGCTCTCAGTAGCTCCAAACAGGGTTCTGGAGCTGTGGGAAAGTTGGGGT 3015
 DB 233 AlaProGlyProAlaGlyAlaArg - - - - - GlySerAspGlySer 245
 QY 3014 AGGGGAAGTTGGGGTAGGGAAATTTTGGCGATGCC - - - - - 2976
 DB 246 ValGlyProValGlyProAlaGlyProIleGlySerAlaGlyProProGlyPheProGly 265
 QY 2975 - - - - - TTCATCAGCCAGTCTCTAGAGAGAGTAGAGGGAGTGGGAAGT 2934
 DB 266 AlaProGlyProLysGlyGluLeuGlyProVal - - - - - GlyAsnProGly 280
 QY 2933 GGGGGGAACCAAGCTGGCCCAAGAGAGGGGTGTTAGGGAAGCCGTGAGACCTGAA 2874
 DB 281 ProAlaGlyProAlaGlyProArgGlyGluAlaGlyLeuProGlyLeuSerGlyProVal 300
 QY 2873 GCGCCACCTCTACCTTCTTCAACACCCCTTAAGCTTGGGTAAACAGCATTTGGAATATCA 2814
 DB 301 GlyProProGly - - - - - AsnProGlyAlaAsnGlyLeuThrGlyAla 314
 QY 2813 TTGGGATAGTAGAATTTCCAGGTCCTGGGTAGGCAATTTGGGGGCCAGACCCAG 2754
 DB 315 LysGlyAlaThrGlyLeuProGlyValAlaGly - - - - - Ala-ProGlyLeuProGlyProAr 333
 QY 2753 GAGAAGAAGATTCTGGCAATGATCAGCCCAATGACCAAGTATCTCAGGGGACCTG - - - - - 2699
 DB 333 g - - - - - GlyIleProGlyProValGlyAl 341
 QY 2698 - - - - - ATTGTTGGGATCCCCACCTACCCCAATATTTAG 2664
 DB 341 aAlaGlyAlaThrGlyProArgGlyLeuValGlyGluProGlyProAlaGlySer - - - - - 359
 QY 2663 ACACCAACACAGAAAAGCTAGCAATGATTCCTTCTACTTGTGTTAAATAAAGTTAA 2604
 DB 360 - - - - - LysGlyGluThrGlyAsnLysGlyGluProGlySer - - - - - 371
 QY 2603 ATATTAAATGCTGTCTCTGTGATGCAACAGAAAGGACCAACAGGCGCACATCTGAT 2544
 DB 372 - - - - - AlaGlyAlaGlnGlyProProGlyProSerGlyGly 383
 QY 2543 AAAAGTAAGAGGGGTGATCAGCAAAAACACAGTGTCT - - - - - GG 2500
 DB 383 uGluGlyLysArgGlySerProGlyGluProGlySerAlaGlyProAlaGlyProProGly 403
 QY 2499 GCTGAGGGA - - - - - CCTGGTTCTTGTGTGTCGCTCAGGACTCTTCCCTTACAATAAG 2443
 DB 403 yLeuArgGlySerProGlySerArgGlyLeuPro - - - - - 414
 QY 2442 TCATATGTTCAATCCCATGGAGGAGTGTTCATCCT - - - - - AGAAACTCCCATGC 2392

415 ----GlyAlaaspGlyArgAlaGlyValMetGlyProProGlyAsnArgGlySerThrG1 433
 QY 2391 AAGAGCTACATTAACAGCAAGCTGCAGGTTAAGGGCTTAGAGATGGGAACACAGGT---- 2336
 DB 433 yProAlaGlyValArgGlyProAsnGlyAspAlaGly-ArgProGlyGluProGlyLeuM 453
 QY 2335 -----GACTCAGTATTATTCAGCTCCCAAAAACCCCTT 2305
 DB 453 etGlyProArgGlyLeuProGlySerProGlyAsnValGlyProAlaGlyLysGluGlyP 473
 QY 2304 CTCTAGGTGTCTCACTACCTAGGAGGTAGCTGTAAACCTTGAGCCCTGGGTAACTCCACCTG 2245
 DB 473 roValGlyLeuProGlyIleAspGlyArgProGlyProIleGlyProAlaGlyProArg 493
 QY 2244 CAGAGTCCCGCATTCAGTGCATGAGCCCTTCTGTCGCTCCCTGTATAGTCCAGACTG 2185
 DB 493 LyGluAlaGlyAsnIleGlyPheProGlyProLysGly----- 505
 QY 2184 AAACCCCTTGGAAAGCCTCCAGTCAGGAGCCCTAGAGACTGGGGAGAGAG----- 2132
 DB 506 -----ProSerGlyAspProGlyLysProGlyLysGlyHisP 519
 QY 2131 -----AGAGGAGCCGCCAGCCCGCTGTCAGCTGTGCAGCTACGACCTCAG 2089
 DB 519 roGlyLeuAlaGlyAlaArgGlyAlaProGlyProAspGly-----AsnAsnG 535
 QY 2088 CAGCAGAGGTGGCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAACTGGCGCCAG 2029
 DB 535 LyAlaGlnGly----- 539
 QY 2028 CCGCGCAGCCCATGGGGCT---NACAGAGCGGGAGCTGGACCCAGTGAAGGAGGCC 1972
 DB 539 roProGlyProGlnGlyValGlnGlyLysGlyGluGlnGlyPro-----AlaGlyP 557
 QY 1971 CTCCA-----CCCCAATGTCTGCGAAGTTTCTACGCTGAGTATTTGGCC 1927
 DB 557 roProGlyPheGlnGlyLeuProGlyProSerGlyThr-----AlaGlyG 572
 QY 1926 AGTGCCTCTTCAATACTACCTGTGTAGCAAAAGTAAATGGCGACAGACCCAGGCT 1867
 DB 572 LuValGly-----Lys--ProGlyG1 578
 QY 1865 GCGGAGACACCATATAGCAGTGACAGCTGCGCTGAGCTGACATGGAGCCCAATAAC 1807
 DB 578 uArgGlyLeuProGly-----GluPheGlyLeuProGlyProAlaGlyProArgG1 595
 QY 1806 AGGGATGGGGCCACTGGGACAGCAGGAGGCACTATCCA-----GGATGGCG 1759
 DB 595 yGluArgGlyProGlyGluSerGlyAlaAlaGlyProSerGlyProIleGlyIleAr 615
 QY 1758 AGGTCCAGCAGATGCCCGCCCGGAACC-----ACCTGGCTCGTGGGCTC 1709
 DB 615 gGlyProSerGlyAlaProGlyProAspGlyAsnLysGlyGluAlaGlyAlaValGly-- 634
 QY 1708 ACCCACCACACGTACGAGACATCAGGAGGCGCCGAGAGGCGCGGGTGGAGG 1649
 DB 635 -----AlaProGlySerAlaGlyAlaLase 642
 QY 1648 TGGGAGCAGCCACTGCTC-----CCAGCACCACGCTGTCCATTAGGAAGGAGGC 1598
 DB 642 rGlyProGlyGlyLeuProGlyGluArgGlyAlaAlaGlyIleProGlyLysGlyG1 662
 QY 1597 TCCAGG-----TTAGG--CCTGGCAGGAGCTGGTGCATCAG 1562
 DB 662 uLysGlyGluThrGlyLeuArgGlyGlyIleGlyAsnProGlyArg-AspGlyAlaArg 682
 QY 1561 GCTGCTCTACTGTAGCAGCTCCAGTGTCCCTCGGTATTTGG----- 1518
 DB 682 LyAlaProGlyAlaIleGlyAlaProGlyProAlaGlyAlaSerGlyAspArgGlyLuA 702
 QY 1517 -----GCAGGAACACCTGCTTCTCCCGGTGGTAGAGGA 1484

[4]
 RN SEQUENCE OF 1-23 FROM N.A.
 RP MEDLINE=87289650; PubMed=3039494;
 RT Rossi P.; de Crombrughe B.;
 RA "Identification of a cell-specific transcriptional enhancer in the
 first intron of the mouse alpha 2 (type I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:5590-5594(1987).
 CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC (FIBRILLAR FORMING COLLAGEN).
 CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 CC HYDROXYAPATITE.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 DR EMBL; X58251; CA941205.1; -;
 DR EMBL; BC007158; AA071158.1; -;
 DR EMBL; K01832; AA37331.1; -;
 DR PIR; A43291; A43291.
 DR MGI; MGI:88468; Colla2.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF01410; COLFI; 1.
 DR ProDom; PD000007; Collagen; 4.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 85
 FT CHAIN 86 1108
 FT PROPEP 1109 1372
 FT MOD_RES 86 86
 FT MOD_RES 90 90
 FT CARBOHYD 1273 1273
 FT CONFLICT 15 15
 FT CONFLICT 1167 1167
 FT SEQUENCE 1372 AA; 129557 MW; 0D17DF5D6C1452D1 CRC64;
 Alignment Scores:
 Pred. No.: 6.42e-07 Length: 1372
 Score: 301.50 Matches: 267
 Percent Similarity: 29.43% Conservative: 42
 Best Local Similarity: 25.43% Mismatches: 357
 Query Match: 4.84% Indels: 385
 DB: 1 Gaps: 52
 US-09-759-143-110 (1-3410) x CA21_MOUSE (1-1372)
 QY 2599 TTAATGCTGCTGCTCTGTGATGGCAACA----- 2570
 Db 12 LeuAlaValThrSerCysLeuAlaThrCysGlnTyrLeuGlnSerGlySerValArg 31
 QY 2569 GAAGGACCAACAGCCACATCTCTGATAAAGGTAAAGGGGGGTGGATCAGCAAAAAGAC 2510
 Db ::::||||| Indels: 4
 Db 32 LysGlyProThrGlyAspArgGlyProArgGlyGlnArgGly----- 45
 QY 2509 AGTGCTGTGGGCTGAGGGGACCTGTCTGTGTGTGTGCTGCTCCCTCAGGACTCTCCCTAC 2450

Db 46 -----ProAlaGly----- 48
 QY 2449 AAATAAGTCATATGTTCAAATCCCATGGAGAGTGTTTCATCTCCTAGAAACTCCCATGCCAA 2390
 Db 48 -----ProAlaGly----- 48
 QY 2389 GAGCTACATTAAAGCAAGCTGCAGTTAAGGGCTTAGAGATGGAAACCAAGCTGACTGA 2330
 Db 49 -----ProArgGly---ArgAspGlyValAspGly----- 57
 QY 2329 GTTTATTACGCTCCCAAAACCCTTCTCTAGGTGTCTCACTAGAGGCTAGCTGTTA 2270
 Db 58 -----ProMetGlyProPro----- 62
 QY 2269 ACCCTGAGCTGGGTATATCCACCTGCAGAGTCCCGCATTCCTCAGTGCATGAGCCCTTCT 2210
 Db 63 -----GlyProGlySerProGlyProGlySerPro---AlaProPro 77
 QY 2209 GGCTCTCCTGTATAAGTCCAGACTGAACCCCTTGAAGGCTCCAGTCAGGC----- 2156
 Db 78 GlyLeuThrGlyAsnPheAlaAlaGlnTyrSerAspLysGlyValSerSerGlyProGly 97
 QY 2155 -----AGCCCTAGAGACTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2111
 Db 98 ProMetGlyLeuMetGlyProArgGlyProProGlyAlaValGlyAlaProGlyProGln 117
 QY 2110 GCTGTGAGCTACGCACCTCAGCAGCAGCAGGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2051
 Db 118 GlyPheGln-----GlyProAlaGlyGluProGlyGluProGlyGlnThrGly 133
 QY 2050 AGCAACAGAAACTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1991
 Db 134 ProAlaGlyProArgGlyProAlaGlySerProGlyLysAlaGlyGluAspGlyHisPro 153
 QY 1990 GGA-----CCCAGTGAGGAGCAGC-----CCTCCACCCCAATGTCTGGAAGTTT 1946
 Db 154 GlyLysProGlyArgProGlyGluArgGlyValGlyProGlnGlyAlaArgGlyPhe 173
 QY 1945 CTACGCTGAGTATTGGCCCAAGTCGCTTTGCAATATACCTGTGTGAAGAAATAAT 1886
 Db 174 Pro-----GlyThrProGlyLeuProGlyPhe-----LysGlyVallys 186
 QY 1885 GCGCAGCAGACCCAGGCGCTGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1835
 Db 187 Gly-----HisSerGlyMetAspGlyLeuLysGlyGln 197
 QY 1834 -----GCTGAGCTGGACATGGAGCCCAATAAACACAGGATGGGCCACCTG 1790
 Db 198 ProGlyAlaGlnGlyValLysGlyGluProGlyAlaProGlyGluAsn---GlyThrProG 217
 QY 1789 GGACAGCAGGA---AGGCACCTATCCAGGATGCGAGGT-----CCAGGCAGAT 1745
 Db 217 LysAlaGlyAlaArgGlyLeuProGlyGluArgGlyValGlyAlaProGly-Pro 236
 QY 1744 GCCCGCGCGGGAACCCCTGGCTGGTGGCTGCCTCCACCCACCCACCCACCCACCCACCCACCC 1685
 Db 237 AlaGlyAlaArgGlySerAspGly-SerValGly---ProValGlyProAlaGlyProIl 255
 QY 1684 ATCAGAGGAGAGGCGCGCAGAGCGCGGAGGTGGAGTGGAGCAGG----- 1639
 Db 255 eGlySerAlaGlyProProGlyPheProGlyAlaProGlyProLysGlyGlyLeuGlyPro 275
 QY 1638 -----CCACTGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1592
 Db 275 oValGlyAsnProGlyProAlaGlyProAlaGlyProArgGlyGlyValGlyLeuProG 295
 QY 1591 CTTA---GGGCTTGGCAGGAAGAGTGGTCTCAGGCTGTCTCAGTGTAGCAGCTCCAGT 1535
 Db 295 yLeuSerGlyPro-----ValGlyProProG 304
 QY 1534 GTCCCTCGGTATT-----TGGG 1517
 ::::|||||

Db 304 yAsnPro-GlyThrAsnGlyLeuThrGlyAlaLysGlyAlaThrGlyLeuProGlyVala 324
 QY 1516 CAGGAACACCTGCTTCTCCGGTGGTACAGGAGGCCA-----GTGTGTA 1472
 Db 324 laGlyAlaProGlyLeuProGlyProArgGlyIleProGlyProAlaGlyAlaAlaGlyA 344
 QY 1471 GGGCAGGATCTCAGCGCTGAGAGGTGCAACCCGGTGAGGCGCTGAAGCTGCACCAC 1412
 Db 344 laThrGlyAlaArgGlyLeu-ValGlyLeuProGlyProAlaGlySerLysGlyLysSer 363
 QY 1411 GGGCAC-----ACTGTGGCAGGAGCTGTGGCAGCGGACGACGACGAGGAA 1367
 Db 364 GlyAsnLysGlyLeuProGlySerValGlyAlaGlnGlyProProGlyProSerGlyGlu 383
 QY 1366 AGCTGCCACCTGGCCAAATAGACTGCTCGAGTGCCGAATC-----GCTGCACGACC 1314
 Db 384 GluGlyLysArgGly-----SerProGlyGluAlaGlySerAlaGlyProAla 399
 QY 1313 GCTCCATGACGAGAGAGACGAGGAGATGGCGCACTGCAGGAACACGACCCGAGGCTGC 1254
 Db 400 GlyPro-----ProGlyLeuArgGlySerProGlySerArg---GlyLeu 413
 QY 1253 CCATCCGAACGCTTCATCATATAGTGTCTCCGGGCT-----CGGTGC 1212
 Db 414 ProGlyAlaAspGlyArgAlaGlyValMetGlyProProGlyAsnArgGlySerThrGly 433
 QY 1211 CGGGTCACCTGGGCAGCGCTGTACAGCCCTCGCCACGAAATCCGTGTAACAA 1152
 Db 434 ProAlaGlyIleArgGlyProAsnGlyAspAlaGlyArgProGlyGlyLeuMet 453
 QY 1151 GCGTGAAGTCATAGTGCATCCAGCTGCACAGCTCAGCCAGGAGGCGGCGCAGGG 1092
 Db 454 GlyProArgGlyLeuProGlySerProGlyAsnValGlyProSerGlyLysGlyGlyPro 473
 QY 1091 TGGGGGCTATGGGC-----AGCACACCTGGTGCACCGGG 1056
 Db 474 ValGlyLeuProGlyIleAspGlyArgProGlyProIleGlyProAlaGlyProArgGly 493
 QY 1055 GAAGCAGGCGCCAGGTTCGGAAAGCCAGCGGCGCGCATGGACAGCAGTGGGGCG 996
 Db 494 GluAlaGlyAsnIleGlyPheProGlyProLysGlyProSerGlyAspProGlyLysPro 513
 QY 995 ACAAGGAGGGCGGACAGCCCTTCGTGCTGGTGGGCGGCGCAGCGTG-----945
 Db 514 GlyGluArgGlyHisProGlyLeuAlaGlyAlaArgGlyAlaProGlyProAspGlyAsn 533
 QY 944 -----CCTCCTCAGCCAGCAGCAGTGTGCTGTAGCAGGAGTGAAGA 900
 Db 534 AsnGlyAlaGlnGlyProProGlyProGlnGlyValGln-----546
 QY 899 TGAGGCTGAGCAGGCGCAAGAGGCACT-----CCTCCTGGGTGGCCAGGTAGGGGGCCA 846
 Db 547 ---GlyGlyLysGlyGluGlnGlyProAlaGlyProProGlyPheGlnGly-----562
 QY 845 GGGCAGTGGTGTCCAGCTCAATGGCAGGAGGAGTGTAGCCAGGAGCCCGCCAGACTGA 786
 Db 563 -----LeuProGlyProSerGlyThrGlyGluValGlyLysProGlyGlu-----578
 QY 785 TCATGAAGCATAGACAGATAGGCTTGGGACAGCTGGT-----CCGGGTCCCGGAGAGGT 729
 Db 579 -----ArgGlyLeuProGlyGlyPheGlyLeuProGlyPro-----590
 QY 728 CAGAGACGAGGCGCTCCAGTGGAGTGAAGCACACCTGGCCACAGAGTCCAGAGCCCA 669
 Db 591 -----AlaGlyProArgGlyGluArgGlyThrProGlyLysGlyAlaAlaGlyPro 608
 QY 668 -----CGCCAGGATGAGCAGTGCAGCTCA-----642
 Db 609 SerGlyProIleGlySerArgGlyProSerGlyAlaProGlyProAspGlyAsnLysGly 628
 QY 641 -----GGGGCTGGGATCCGGGCACAGCAGCCCTGTGTAGCCAGCCGCGCC 597
 Db 629 GluAlaGlyAlaValGlyAla-----ProGlySerAlaGlyAlaSerGlyProGlyGly 646

QY 596 TTGGGATGAGAAAGAGCTCAGCAGATGCCCA-----564
 Db 647 LeuProGlyGluArgGlyAlaAlaGlyIleProGlyGlyLysGlyGluLysGlyGluThr 666
 QY 563 -----AGGACAGTCCAGATCA 546
 Db 667 GlyLeuArgGlyAspThrGlyAsnThrGlyArgAspGlyAlaArgGlyIleProGly-----685
 QY 545 AGGGCGGCGGCGCCATACGCTCCAGCGCAGTGGTCACTGGCTGAGCCTAGGAGCGGGA 486
 Db 686 AlaValGlyAlaProGlyProAlaGlyAlaSerGlyAspArgGlyGluAlaGlyAlaAla 705
 QY 485 CACAGACCCAGCCAGCAGTGGACCAATGCCAGCACCACCTGCTCATGAACCTCTCTCTA 426
 Db 706 GlyProSerGlyProAla-----711
 QY 425 CCCCACCTCCAGCAGCAGGCGGCACATAGGTGCTGCGCGCAACACACACCTCCA 366
 Db 712 -----GlyProArgGlySerPro 717
 QY 365 GGCCAAAG-----TTAGCAGGTTCAGCAGCAAGA 336
 Db 718 GlyGluArgGlyGluValGlyProAlaGlyProAsnGlyPheAlaGlyProAlaGlyAla 737
 QY 335 GCTGGCTTTCCGGTGCC-----318
 Db 738 AlaGlyGlnProGlyAlaLysGlyGlyLysGlyThrLysGlyProLysGlyGluAsnGly 757
 QY 317 -----CGACGAGCGCTCACCACAGCCTCTGGACCATAGTGGGCGCAGCGCG-----269
 Db 758 IleValGlyProThrGlySerValGlyAlaAlaGlyPro-SerGlyProAsnGlyProPr 777
 QY 268 -----TAGGCTC 261
 Db 777 oGlyProValGlySerArgGlyAspGlyGlyProProGlyMetThrGlyPheProGlyAl 797
 QY 260 AGGGGCGGTTCAGGCACTCCAGAACTGCTGCTCGGCTCTCCTCCAGAACTGCGGC 201
 Db 797 aAlaGlyArgThrGlyProPro-----cIyProSerGlyIleAl 810
 QY 200 CTCCTCTCTGCTGCTGCCCAACTGCCTAGGAATCAGCCAGCGCCCATTTTCTCCAGCC 141
 Db 810 aGlyProPro-----GlyProProGlyAla-----818
 QY 140 CTTTGGTCCCGTCCAGCTTCTCAGCCCATCTCAACACCTGCTGCTGGGCGCACCTCA 81
 Db 819 -AlaGly-LysGluGlyIleArgGlyProArgGlyAspGlnGlyProValGlyArg---t 837
 QY 80 GTGGGACACGCTCTCATCTCAGATCCTGCG-----CGAGGCGCGGCTGT 33
 Db 837 hrGlyGluThrGlyAlaSerGlyProGlyPheValGlyGlyLysGlyProSerGlyG 857
 QY 32 CACCGGAGCGCAGC 19
 Db 857 LuProGlyThrAla 861

RESULT 41

CA2B_HUMAN

ID CA2B_HUMAN

STANDARD; PRT: 1736 AA.

AC P13942; Q13273; Q13271; Q13272; Q0751; Q99866; Q9UIP9;

DT 01-JAN-1990 (Rel. 13, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Collagen alpha 2(XI) chain precursor.

GN COL11A2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homiq.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96032717; PubMed=7559422;

RA Vuoristo M.M., Pihlajamaa T., Vandenbergh P., Prockop D.J.,
 RA Ala-Kokko L.;
 RT "The human COL1A2 gene structure indicates that the gene has not
 RT evolved with the genes for the major fibrillar collagens.";
 RL J. Biol. Chem. 270:22873-22881(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Tubby B.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 59-807 FROM N.A.
 RN [3]
 RP TISSUE-Cartilage;
 RC MEDLINE=93314796; PubMed=8325374;
 RA Zhidkova N.I., Brewton R.G., Wayne R.;
 RT "Molecular cloning of PARP (proline/arginine-rich protein) from human
 RT cartilage and subsequent demonstration that PARP is a fragment of the
 RT NH2-terminal domain of the collagen alpha 2(XI) chain.";
 RL FEBS Lett. 326:25-28(1993).
 RN [4]
 RP SEQUENCE OF 730-1690 FROM N.A.
 RX MEDLINE=89340485; PubMed=2760050;
 RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,
 RA van der Rest M., Ono K., Solomon E., Ninomiya Y., Olsen B.R.;
 RT "The human alpha 2(XI) collagen (COL1A2) chain. Molecular cloning of
 RT cDNA and genomic DNA reveals characteristics of a fibrillar collagen
 RT with differences in genomic organization.";
 RL J. Biol. Chem. 264:13910-13916(1989).
 RN [5]
 RP SEQUENCE OF 1-537 FROM N.A.
 RX MEDLINE=96435918; PubMed=8838804;
 RA Lui V.C., Ng L.J., Sat E.W., Cheah K.S.;
 RT "The human alpha 2(XI) collagen gene (COL1A2): completion of coding
 RT information, identification of the promoter sequence, and precise
 RT localization within the major histocompatibility complex reveal
 RT overlap with the KES gene.";
 RL Genomics 32:401-412(1996).
 RN [6]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=95238468; PubMed=7721876;
 RA Zhidkova N.I., Justice S.K., Wayne R.;
 RT "Alternative mRNA processing occurs in the variable region of the
 RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";
 RL J. Biol. Chem. 270:9486-9493(1995).
 RN [7]
 RP DISEASE.
 RX PubMed=10677296;
 RA Melkonian M., Brunner H.G., Manouvier S., Hennekam R.,
 RA Superti-Furga A., Kaeerlaeinen H., Pauli R.M., van Essen T.,
 RA Warman M.L., Bonaventure J., Miny P., Ala-Kokko L.;
 RT "Autosomal recessive disorder otospondyloymeagapiphyseal dysplasia is
 RT associated with loss-of-function mutations in the COL1A2 gene.";
 RL Am. J. Hum. Genet. 66:368-377(2000).
 RN [8]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [9]
 RP VARIANT OSMED ARG-661.
 RX MEDLINE=95163096; PubMed=7859284;
 RA Vikkula M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,
 RA Goldring M.B., van Beersum S.E.C., de Waal Malefijt M.C.,
 RA van den Hoogen F.H.J., Ropers H.-H., Wayne R., Cheah K.S.E.,
 RA Olsen B.R., Warman M.L., Brunner H.G.;
 RT "Autosomal dominant and recessive osteochondrodysplasias associated
 RT with the COL1A2 locus.";
 RL Cell 80:431-437(1995).
 RN [10]
 RP VARIANTS GLY-593; LYS-824; LEU-879; THR-1316 AND GLN-1600.
 RX PubMed=9585596;

RA Koga H., Sakou T., Taketomi E., Hayashi K., Numasawa T., Harata S.,
 RA Yone K., Matsunaga S., Otterud B., Inoue I., Leppert M.;
 RT "Genetic mapping of ossification of the posterior longitudinal
 RT ligament of the spine.";
 RL Am. J. Hum. Genet. 62:1460-1467(1998).
 RN [11]
 RP VARIANT WZS GLU-955.
 RX PubMed=9805126;
 RA Pihlajamaa T., Prockop D.J., Faber J., Winterpacht A., Zabel B.,
 RA Giedion A., Wiesbauer P., Spranger J., Ala-Kokko L.;
 RT "Heterozygous glycine substitution in the COL1A2 gene in the original
 RT patient with the Weissenbacher-Zweymueller syndrome demonstrates its
 RT identity with heterozygous OSMED (nonocular Stickler syndrome).";
 RL Am. J. Med. Genet. 80:115-120(1998).
 RN [12]
 RP VARIANT STL3 940-GLY--PRO-948 DEL.
 RX PubMed=9506662;
 RA Sirko-Osadsa D.A., Murray M.A., Scott J.A., Lavery M.A., Warman M.L.,
 RA Robin N.H.;
 RT "Stickler syndrome without eye involvement is caused by mutations in
 RT COL1A2, the gene encoding the alpha-2(XI) chain of type XI
 RT collagen.";
 RL J. Pediatr. 132:368-371(1998).
 RN [13]
 RP VARIANTS DFNA13 GLU-808 AND CYS-1034, AND REVISIONS TO 1031-1032.
 RX MEDLINE=20047768; PubMed=10581026;
 RA McGuire W.T., Prasad S.D., Griffith A.J., Kunst H.P.M., Green G.E.,
 RA Shpargel K.B., Runge C., Huybrechts C., Mueller R.F., Lynch E.,
 RA King M.-C., Brunner H.G., Cremers C.W.R.J., Takanosu M., Li S.-W.,
 RA Arita M., Wayne R., Prockop D.J., Van Camp G., Smith R.J.H.;
 RT "Mutations in COL1A2 cause non-syndromic hearing loss (DFNA13).";
 RL Nat. Genet. 23:413-419(1999).
 CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.
 CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
 CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD
 CC OF ALPHA 3(XI)-1(II).
 CC -!- ALTERNATIVE PRODUCTS: 8 isoforms; 1 (shown here), 2, 3, 4, 5, 6, 7
 CC and 8; may be produced by alternative splicing. They lack exons 6,
 CC 7 or 8 or a combination of these exons.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- PTM: A DISULFIDE-BONDED PEPTIDE CALLED PROLINE/ARGININE-RICH
 CC PROTEIN OR PARP IS RELEASED FROM THE AMINO TERMINUS DURING
 CC EXTRACELLULAR PROCESSING AND IS SUBSEQUENTLY RETAINED IN THE
 CC CARTILAGE MATRIX FROM WHICH IT CAN BE ISOLATED IN SIGNIFICANT
 CC AMOUNTS.
 CC -!- DISEASE: Defects in COL1A2 are the cause of Stickler syndrome
 CC type 3 (STL3). It is an autosomal dominant disorder characterized
 CC by oro-facial, auditory and skeletal manifestations, such as
 CC midfacial hypoplasia, cleft palate, osteoarthritis, and
 CC sensorineural hearing loss. Differently from Stickler syndrome
 CC type 1 and 2, no ocular involvement is observed. This disorder is
 CC also referred to as Stickler-like syndrome or non-ocular Stickler
 CC syndrome.
 CC -!- DISEASE: Defects in COL1A2 are the cause of autosomal recessive
 CC otospondyloymeagapiphyseal dysplasia (OSMED), a skeletal dysplasia
 CC accompanied by severe hearing loss. The phenotype overlaps that of
 CC autosomal dominant skeletal disorders (Stickler and Marshall
 CC syndromes) but can be distinguished by disproportionately short
 CC limbs and lack of ocular involvement.
 CC -!- DISEASE: Defects in COL1A2 are the cause of Weissenbacher-
 CC zweymueller syndrome (WZS), an autosomal dominant disorder allelic
 CC with STL3 and OSMED. It is also referred to as heterozygous OSMED.
 CC -!- DISEASE: Defects in COL1A2 are the cause of autosomal dominant
 CC nonsyndromic sensorineural deafness type 13 (DFNA13). Affected
 CC individuals experience progressive hearing loss beginning in the
 CC second to fourth decades, eventually making use of amplification
 CC mandatory.
 CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
 CC -!- DATABASE: NAME-Hereditary hearing loss homepage;
 CC NOTE=Gene page;

QY	1786	CAGCAGGAGGCACTATCCA	---GGATGGCAGGTCTCCAGGCAGATGCCCGCCGGCAA	1731
Db	932	laAlaIaGluThrGlyProMetGlyGluArgGlyHis	-----ProGlyProProG 949	
QY	1730	--CCACCTGGCGTGGCTCACC	---ACCACACAGTACGAGACATCACAGGCAG	1673
Db	949	lyPro-ProGlyGluGlnGlyLeu	-ProGlyThrAlaGlyLysGluGlyThrLysGlyAs	968
QY	1672	G	-----GCCCGCAGACGCGGTGGAGTGGGAGCAGCCACTGCCTCC	1628
Db	968	pProGlyProProGlyAlaProGlyLysAspGlyProAlaGlyLeuArgGlyPheProGly	-----1588	
QY	1627	AGCACCCAGTGTCCATTAGGAAGGAGTCCAGGCTTA	-----1588	
Db	988	yGluArgGlyLeuProGlyThrAlaGlyGlyProGlyLeuLysGlyAsnGluGlyProse	1008	
QY	1587	-----GGGCTGCGCAAGC	-----TGGTCATCAGGCTGTC	1556
Db	1008	rGlyProProGlyPro-AlaGlySerProGlyGluArgGlyAlaAlaGlySerGlyGlyP	1028	
QY	1555	CTCACTGCTAGCACTCCAGTGTCCCTCGGTATTGGGCAGGAACACCTGCTCTCCCG	1496	
Db	1028	ro	-----11eGlyProProGlyArgProG 1036	
QY	1495	GTGGTAGAGGAGGCCAGTGTGTAGGCCAGGATCTGCAGGCTGAGAAGGTGAACCGGT	1436	
Db	1036	lyProGlnGlyProPro	-----GlyAlaAlaGlyGluLys-GlyValProGly 1051	
QY	1435	GAGGGCGGTGAAGCTGTCA	---ACCACGCCACACTGTGGCAGCAGCATGTGCACCGGCAGC	1376
Db	1052	GluLysGly--ProIleGlyProThrGlyArgAspGlyVal	-----GlnGlyP 1067	
QY	1375	CACAGGAAAGCTGCCACACTGGCCCAATAGACTGCTCAGTGCCGAATCGTGCACCAG	1316	
Db	1067	roValGly	-----LeuProGlyProA 1074	
QY	1315	CCGTCCTCAACAGAGAGAACACCGAGATGGCGCACTGCAGGAACA	-----1266	
Db	1074	laGlyPro	-----ProGly-----ValAlaGlyLysAspGlyAspL 1086	
QY	1265	-----GCCCGAGGTGCCCATCCGACGCCTTCATCATATGTCCTCGGCG	1220	
Db	1086	ysGlyGluValGlyAspProGlyClnLysGlyThrLysGlyAsnLysGlyGluHisGlyP	1106	
QY	1219	CTCGTCCCGCTCAGCTCTGGGCACGCCTGTGTACAGCCCTCGCCACGAAATCCGT	1160	
Db	1106	ro	-----ProGly--ProProGlyProIleGlyPro--1115	
QY	1159	GTAACACGGTGAAGTTCATGATGCCATCCAGCTGCACACTGCACGCACGACGACGACGAGCGG	1100	
Db	1116	-----ValGlyGlnProGlyAlaAlaG 1123		
QY	1099	GCACAGG	-----TCCGGGGCATGGCGCAGCACAGCT-----GTCGACGCGGGGAAG	1052
Db	1123	lyAlaAspGlyGluProGlyAlaArgGlyProGlnGlyHisPheGlyAlaLysGlyAspG	1143	
QY	1051	CAGGCGCCAGGTTCCGAAAGCCACGCGGCC	---GGCATGTGCACG	1005
Db	1143	luGlyThrArgGlyPheAsnGlyProProGlyProIleGlyLeuGlnGlyLeuProGlyP	1163	
QY	1004	--AGTGGGCGCAAGAGGGGGCGCAGACGCCCTTCTGCTGCTCGGTGGGGCCCGCAGGC	947	
Db	1163	roSerGly	-----GluLysGlyGluThrGlyAspVal-----GlyProMetG 1177	
QY	946	TGCTCTCAGCCACCGACAGTGTGGCTGCTACGCAAGTGAAGAGATGAGGCTGAGCAG	887	
Db	1177	lyProProGlyProPro	-----GlyProArgGlyProAlaG 1189	
QY	886	GCCAAAGAGCACT	-----CCTCTCGGTGCCAGGTAGG---GGGCCAG	845
Db	1189	lyProAsnGlyAlaAspGlyProGlnGlyProProGlyGlyValGlyAsnLeuGlyPro	1208	

QY	844	GGCACTGGTGTCCCACTCAATGCGCAGGAGGAGGTAGCCAGCAGCGCCCAAGACTGAT	785
DB	1209	-----ProGlyGluLysGlyGluProGlyGluSerPro	1221
QY	784	CATGAAGGCATAGACAGACTAGCCCTGGCGGACAGTGGTCCGGGTCCCGGAAGAGGTCAGA	725
DB	1222	-----GlyIleGlnGlyProGlyValLysGlyProArgGlyGlyLuarGlyGlyL	1239
QY	724	GA-----GCAGGGCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAAGTCCACGACGCC	671
DB	1239	yselyGluSerGlyGlnProGlyGluProGlyProGlyAlaLysGlyProGlnGlyA	1259
QY	670	CA-----CGCCAGGATGAGCAGTGCACAGCTCCAGGCGCTCGGATCCCG	636
DB	1259	sPAspGlyProLysGlyAsnProGly-----ProValGlyPheProGly-----AspProG	1276
QY	625	GCACAGACCCCTGCTAGCCAGCCGG-----	600
DB	1276	lyProProGlyGluGlyProArgGlyGlnAspGlyAlaLysGlyAspArgGlyGluA	1296
QY	599	-----C	599
DB	1296	spGlyGluProGlyGlnProGlySerProGlyProThrGlyGluAsnGlyProProGlyP	1316
QY	598	CCTTGGGATGAAAGAGGCTCAGCAGGATGCCCAAGACAGTGCACAGATGAAGGCCG	539
DB	1316	roLeuGly--LysArgGlyProAlaGlySerProGlySerGlyGlyArg-----GlnGlyG	1334
QY	538	GGCGGCGCATAGGCTGCACGCCAGTGGTCACTGGCTGAGCTAGGAGGGGACACAGAC	479
DB	1334	lyLysGly-----AlaLysGlyAspProGlyAlaIleGlyAlaProGlyLysT	1350
QY	478	CAGGCCACGACCTGGACCAATGCCACACCATGGTCATGAACCTCTCTTACCCCCAC	419
DB	1350	hrGlyProValGlyProAlaGlyProAlaGlyLysProGlyProAspGlyLeuArgGlyL	1370
QY	418	TTCCAGCAGCAGAGCGCGCATAGGTGATGCTCGGCGGCCAAACACA-----	372
DB	1370	euProGlySerValGlyGlnGly-----ArgProGlyAlaThrGlyGlnAlaG	1387
QY	371	--CTTCCAGGCAAAAGTTAGCAGGTGACACAGAGCTGGGCTTCCGTTGCCGCGAC	314
DB	1387	lyProProGlyPro-----ValGlyProProGlyLeuProGlyLeuArgGlyAspAlaG	1405
QY	313	CAGGC-----	309
DB	1405	lyAlaLysGlyGluLysGlyHisProGlyLeuIleGlyLeuIleGlyProProGlyGluG	1425
QY	309	-----	309
DB	1425	lnGlyGluLysGlyAspArgGlyLeuProGlyProGlnGlySerProGlyGlnLysGlyG	1445
QY	308	-----GGCTCACCACAGCTCTGGACCATTAGTGGCCAGGCGGG-----	269
DB	1445	luMetGlyIleProGlyAlaSerGlyPro-IleGlyProGlyGlyProProGlyLeuPro	1464
QY	268	-----TAGGGCTCAGGGGCGGTTACGGCACT	242
DB	1465	GlyProAlaGlyProLysGlyAlaLysGlyAlaThrGlyProGlyGlyProLysGlyGlu	1484
QY	241	CGAACTGCTTCTCGGCTCTGCTCCAGAGCTGGCGCTCTCTCTCTCT-----TGC	188
DB	1485	LysGlyVal-----GlnGlyProProGlyHisProGlyProProGlyGluVal	1500
QY	187	TGCCGCAACTGCTTAGGAATCAGCCA-----GGCGCCCAT	152
DB	1501	IleGlnProLeuProIleGlnMetProLysLysThrArgArgSerValAspGlySerArg	1520
QY	151	TTCTGCCAG-----CCCTTGTGGCGGTCACGCTTCTACGCCATGCTC	107
DB	1521	LeuMetGlnGluAspGluAlaIleProThrGlyGlyAlaProGlySerProGlyGlyLeu	1540
QY	106	AACACCTGTGCTGGGCGACCTCAGTGGGACACGCTCTCATCACTCAGACG	56

Db 1541 GluGlutile-----PheGlySerLeuAspSerLeuArg 1551
 RESULT 42
 CA334_HUMAN
 ID CA334_HUMAN STANDARD; PRT; 1670 AA.
 AC Q01955; Q9BQ02;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).
 GN COL4A3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Kidney;
 RC MEDLINE=94364994; PubMed=8083201;
 RA Maruyama M., Leinonen A., Mochizuki T., Tryggvason K., Readers S.T.;
 RT "Complete primary structure of the human alpha 3(IV) collagen chain.
 RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
 RT human tissues.";
 RL J. Biol. Chem. 269:23013-23017(1994).
 RN [2]
 RP REVISIONS.
 RA Leinonen A.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A., VARIANTS AS E-297; R-407; R-640; R-1167; E-1207;
 RP Q-1215; S-1277; T-1330; E-1334; E-1347 AND C-1661, AND VARIANTS R-43;
 RP E-162; Y-326; H-408; R-451; L-574; E-1269 AND P-1474.
 RX MEDLINE=21064696; PubMed=11134255;
 RA Heidt L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
 RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
 RA "Structure of the human type IV collagen gene COL4A3 and mutations in
 RT autosomal Alport syndrome.";
 RL J. Am. Soc. Nephrol. 12:97-106(2001).
 RN [4]
 RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=93015826; PubMed=1400291;
 RA Quinones S., Bernal D., Garcia-Soto M., Elena S.F., Saus J.;
 RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
 RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
 RT antigenic region at the triple helix/NC1 domain junction.";
 RL J. Biol. Chem. 267:19780-19784(1992).
 RN [5]
 RP SEQUENCE OF 1453-1670 FROM N.A.
 RX MEDLINE=91353570; PubMed=1882840;
 RA Morrison K.E., Maruyama M., Yang-Feng T.L., Readers S.T.;
 RT "Sequence and localization of a partial cDNA encoding the human alpha
 RT 3 chain of type IV collagen.";
 RL Am. J. Hum. Genet. 49:545-554(1991).
 RN [6]
 RP SEQUENCE OF 1331-1670 FROM N.A.
 RX MEDLINE=92147878; PubMed=1737849;
 RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,
 RA Pusey C.D.;
 RT "Molecular cloning of the human Goodpasture antigen demonstrates it
 RT to be the alpha 3 chain of type IV collagen.";
 RL J. Clin. Invest. 89:592-601(1992).
 RN [7]
 RP SEQUENCE OF 1644-1670 FROM N.A.
 RX TISSUE=Kidney;
 RA Ding J.;
 RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
 RC TISSUE=Kidney;
 RX MEDLINE=94124597; PubMed=8294492;
 RA Feng L., Xia Y., Wilson C.B.;

"Alternative splicing of the NCI domain of the human alpha 3(IV) collagen gene. Differential expression of mRNA transcripts that predict three protein variants with distinct carboxyl regions.";
 J. Biol. Chem. 269:2342-2348(1994).
 [9]
 RN SEQUENCE OF 1-29 FROM N.A.
 RP MEDLINE=98196854; PubMed=9537506;
 RX Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshiooka H.,
 RA Ninomiya Y.;
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and alpha4(IV) collagen chains are arranged head-to-head on chromosome 2q36.";
 RL FEBS Lett. 424:11-16(1998).
 RN [10]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=93280184; PubMed=8505332;
 RA Bernal D., Quinones S., Saus J.;
 RT "The human mRNA encoding the Goodpasture antigen is alternatively spliced.";
 RL J. Biol. Chem. 268:12090-12094(1993).
 RN [11]
 RP VARIANT PRO-1474.
 RX MEDLINE=95078827; PubMed=7987301;
 RA Lemink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
 RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
 RA Readers S.T., Smeets H.J.M.;
 RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal recessive Alport syndrome.";
 RL Hum. Mol. Genet. 3:1269-1273(1994).
 CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/NIDOGEN.
 CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2/V AND 3/L5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR C-TERMINAL NCI DOMAINS.
 CC -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE COLocalized AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE, COCHLEA, LUNG AND BRAIN.
 CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG-CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL N-LINKED GLYCOSYLATION SITE.
 CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.
 CC -!- PTM: Phosphorylated by the Goodpasture antigen-binding protein.
 CC -!- DISEASE: ANTIBODIES AGAINST THE NCI DOMAIN OF ALPHA3(IV) MEDIATE THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
 CC -!- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE, HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN MALES AND FEMALES.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

CC	EMBL; X80031; CAA56335.1; -	QY	3008	AAATTGGGGTAGGGGAAATTTGGCAGTCCTTCATCAGCCAGTCCTAGAGAGTA	2949
DR	EMBL; AJ288487; CAC36101.1; JOINED.	Db	92	ValArgGlyIleSerGlyLeuProGlyPheSerGly	110
DR	EMBL; AJ288488; CAC36101.1; JOINED.	QY	2948	GAGGGG-	2928
DR	EMBL; AJ288489; CAC36101.1; JOINED.	Db	111	ProGlyAsnThrGlyProTyrglyLeuValGlyValProGlyCysSerGlySerGly	130
DR	EMBL; AJ288490; CAC36101.1; JOINED.	QY	2927	AACCAAGGCTGGGCCAAGAGAAGAGGGTGGTGGGAAGCCCTGGAGACCTGAAGCCCA	2868
DR	EMBL; AJ288491; CAC36101.1; JOINED.	Db	131	GlulnGlyPhePro-	147
DR	EMBL; AJ288492; CAC36101.1; JOINED.	QY	2867	CCCTCTACCTTCCTTCAA-	2820
DR	EMBL; AJ288493; CAC36101.1; JOINED.	Db	148	GlyAlaAlaGlyLeuLysGlyGlnLysGlyAlaPro-AlaLysGlyGluAspIleGluLe	167
DR	EMBL; AJ288494; CAC36101.1; JOINED.	QY	2819	TTATCATTTGGGATGAGTAGAATTTCCAGGTCTCTGGGTAGGCAATTTGGGGCCAGA	2760
DR	EMBL; AJ288495; CAC36101.1; JOINED.	Db	167	uAspAlaLysGlyAspProGlyLeuPro-	180
DR	EMBL; AJ288496; CAC36101.1; JOINED.	QY	2759	CCCCAGGAGAAGAAGATTCTGGCAATGATCAGCCCAATGACCACTATCTCAGGGACCT	2700
DR	EMBL; AJ288497; CAC36101.1; JOINED.	Db	180	yPro-	193
DR	EMBL; AJ288498; CAC36101.1; JOINED.	QY	2699	GATTGTTGGGATCCCCACCCCTACCCAAATATTTAGACACCAACACAGAAAGCTAGCAA	2640
DR	EMBL; AJ288499; CAC36101.1; JOINED.	Db	194	ValGlyProProGlyPro-	203
DR	EMBL; AJ288500; CAC36101.1; JOINED.	QY	2639	TGGATTCCTCTACTTTGTTTAATAATAAGTTAAATATTTAAATGCCGTGCTCTGT	2580
DR	EMBL; AJ288501; CAC36101.1; JOINED.	Db	203	eGlyPhePro-	208
DR	EMBL; AJ288502; CAC36101.1; JOINED.	QY	2579	GATGGCAACAGAAGGACCAACAGGCCAC-	2535
DR	EMBL; AJ288503; CAC36101.1; JOINED.	Db	208	aMetGlyProArgGlyProLysGlyHisMetGlyGluArgValIleGlyHisGlyGly	228
DR	EMBL; AJ288504; CAC36101.1; JOINED.	QY	2534	GAGGGGGTG-	2525
DR	EMBL; AJ288505; CAC36101.1; JOINED.	Db	228	uArgGlyValLysGlyLeuThrGlyProProGlyProGlyThrValIleValThrLe	248
DR	EMBL; AJ288506; CAC36101.1; JOINED.	QY	2524	GATCAGCAAAAGACAGTGTGTGGGTGAGGGGACCTGTTCTTGTG-	2477
DR	EMBL; AJ288507; CAC36101.1; JOINED.	Db	248	uThrGlyProAspAsnArgThrAspLeuLysGlyGluLysGlyAspGlyAlaMetGly	268
DR	EMBL; AJ288508; CAC36101.1; JOINED.	QY	2476	TGTGGCCCTCAGGACTCTTCCCTCAAAATAAGTCATATGTTCAAA	2430
DR	EMBL; AJ288509; CAC36101.1; JOINED.	Db	268	yGluProGlyProProGlyProSerGlyLeu-	278
DR	EMBL; AJ288510; CAC36101.1; JOINED.	QY	2429	TCCCATGGAGGAGTGTTCATCCTCAGAACTCCCATGCAAGACTACATTAACGAGCT	2370
DR	EMBL; AJ288511; CAC36101.1; JOINED.	Db	279	ProGlyGluSer-	282
DR	EMBL; AJ288512; CAC36101.1; JOINED.	QY	2369	GCAGGTTAAGGGCTTAGAGATGGGAACACAGGTGACTGAGTTATTATTCAGTCCCAAAA	2310
DR	EMBL; AJ288513; CAC36101.1; JOINED.	Db	283	TyrGlySerGluLysGlyAlaProGlyAsp-	292
DR	EMBL; AJ288514; CAC36101.1; JOINED.	QY	2309	CCCTTCTCTA-	2271
DR	EMBL; AJ288515; CAC36101.1; JOINED.	Db	293	ProGlyLeuGlnGlyLysProGlyLysAspGlyValPro-	305
DR	EMBL; AJ288516; CAC36101.1; JOINED.	QY	2270	AACCTGTAGCCCTGGGTAAATCCACCTGCAGAGTCCCGCATTCCTAGTGGAGCCCTTC	2211
DR	EMBL; AJ288517; CAC36101.1; JOINED.	Db	306	GlyPheProGlySerGlyGluLysGlyValLysGlyAsnArgGlyPheProGly	321
DR	EMBL; AJ288518; CAC36101.1; JOINED.	QY	2210	TGGCCTCCCTGTATAGTCCACAGCTGAAACCCCTTGAAGCCCTCCAGTCAGG-	2156
DR	EMBL; AJ288519; CAC36101.1; JOINED.	Db	321	y-	335
DR	EMBL; AJ288520; CAC36101.1; JOINED.	QY	2155	AGCCCTAGACACTGGGGAGAGG-	2132
DR	EMBL; AJ288521; CAC36101.1; JOINED.	Db	335	eGlyProGlyPheArgGlyProThrGlyTyrThrAspThrThrGlnGluLysGlyAs	355

Alignment Scores:

Pred. No.:	6.97e-07	Length:	1670
Score:	300.50	Matches:	325
Percent Similarity:	28.31%	Conservative:	60
Best Local Similarity:	23.90%	Mismatches:	396
Query Match:	4.83%	Indels:	579
DB:	1	Gaps:	78

US-09-759-143-110 (1-3410) x CA34_HUMAN (1-1670)

QY 3068 GGTCTCAGTAGTCTCAACAGGTTGTGGAGCTGGTGGGAAAGTGGGGTAGGGGA 3009
 Db 76 GlyProLysGlyPheProGly-----LeuProGlyLeuThrGlySerLysGly 91


```

Db 946 GlutSerHisValIleGlyAspLysGlyGluProGlyLeuLysGlyPheAlaGlyAsn 965
Qy 479 CCAGGC-----CCAGCACTGGACCAA 459
Db 966 ProGlyGlyLysGlyAsnArgGlyValProGlyMetProGlyLeuLysGlyGly 985
Qy 458 TGCCAGCACCATGGTCATGAACCTTCTCTCTCCCTCCCTCCAGCAGCAGAGCGGCA 399
Db 986 LeuProGlyProAlaGly-----ProProGlyProArgGlyAspLeuGly 1000
Qy 398 CATAGGTGATGCTCGCGGCAACACACCT-----369
Db 1001 Ser-----ThrGlyAsnProGlyGluProGlyLeuArgGlyIleProGlySerMetGly 1018
Qy 368 -----CCAGGCCAAAGG-----TTAGCAGGTGTACACACAGCT 333
Db 1019 AsnMetGlyMetProGlySerLysGlyLysArgGlyThrLeuGlyPheProGlyArgAla 1038
Qy 332 GGCCTTCCGGTGCC-----318
Db 1039 GlyArgProGlyLeuProGlyIleHisGlyLeuGlnGlyAspLysGlyGluProGlyTyr 1058
Qy 317 GCAGCAGCGGCTCACCCACAGCCTCTGCACCATAGTGG-----CCAGGGCG 270
Db 1059 SerGluGlyThrArgProGlyProProGlyProProGlyAspProGlyLeuProGlyAs 1078
Qy 269 GTAGGCTCAGGGGCGCTTCAGGCACCTCCA---GAACCTGCTCTCGCTCGCTCTGCT-- 215
Db 1078 pMetGlyLysGlyGluMetGlyGlnProGlyProGlyHisLeuGlyProAlaGly 1098
Qy 214 -CCAGAGCTCGGGCTCTCTCTCTGCTGCGCGCCACTGCTAGATCAGCCAGCGC 156
Db 1098 yProGluGlyAlaProGlySerProGlySerProGlyLeuPro---GlyLysProGlyPr 1117
Qy 155 CCATTCTGCCAGCCCTTT-----GTGCGGCTGCCAGCTCTCA 117
Db 1117 oHisGlyAspLeuGlyPheLysGlyIleLysGlyLeuLeuGly-ProProGlyIleArgg 1137
Qy 116 GCCCATGCTCACACCTGCTGCTGCGGCGAC-----CTCAGTGGGGACA 72
Db 1137 yProProGlyLeuProGlyPheProGlySerProGlyProMetGlyIleArgGlyAspG 1157
Qy 71 CGTCATCATCTCAGATCTGCTG-----49
Db 1157 lnglyArgAspGlyIleProGlyProAlaGlyGlyLysGlyGluThrGlyLeuLeuArga 1177
Qy 48 -----CGAGGCGCGGCTGTACCCGAGCGCGCTGCAGG 10
Db 1177 laProProGlyProArgGlyAsnProGlyAlaGlnGlyAlaLys 1191
RESULT 43
CALB_MOUSE
ID CALB_MOUSE STANDARD; PRT: 1804 AA.
AC Q61245; Q64047;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 35, Last sequence update)
DE Collin1
GE Collagen alpha 1(XI) chain precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=96015067; PubMed=8530046;
RA Yoshioka H., Inoguchi K., Khaleduzzaman M., Ninomiya Y.,
RA Andrikopoulos K., Ramirez F.;
RT "Coding sequence and alternative splicing of the mouse alpha 1(XI)
RL collagen gene (Coll1a1).";
RL Genomics 28:337-340(1995).
RN [2]

```

```

RP SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.
RC STRAIN=C57BL/6;
RX MEDLINE=95163095; PubMed=7859283;
RA Li Y., Lacerda D.A., Warman M.L., Beier D.R., Yoshioka H.,
RA Ninomiya Y., Oxford J.T., Morris N.P., Andrikopoulos K.,
RA Ramirez F., Wardell B.B., Lifferth G.D., Teuscher C., Woodward S.R.,
RA Taylor B.A., Seegmiller R.E., Olsen B.R.;
RT "A fibrillar collagen gene, Coll1a1, is essential for skeletal
morphogenesis.";
RL Cell 80:423-430(1995).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
MODIFICATION OF ALPHA 1(XI). ALPHA 1(XI) CAN ALSO BE FOUND INSTEAD
OF ALPHA 3(XI)=1(XI) (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- DISEASE: DEFECTS IN COL1A1 ARE ASSOCIATED WITH CHONDRODYSPLASIA,
AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY SKELETAL DEFECTS
CAUSED BY ABNORMALITIES IN THE CARTILAGE OF LIMBS, RIBS, MANDIBLES
AND TRACHEA.
CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC -!- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
EMBL; D38162; BAA07367.1; -;
EMBL; S74574; AAB33439.1; -;
MGD; MGI:88446; Coll1a1.
InterPro; IPR000087; Collagen.
InterPro; IPR000885; Fib_collagen_C.
InterPro; IPR001791; Laminin_G.
InterPro; IPR001230; Prenyl_site.
InterPro; IPR003129; TSPN.
Pfam; PF01391; Collagen; 16.
Pfam; PF01410; COLFI; 1.
Pfam; PF02210; TSPN; 1.
ProDom; PD000007; Collagen; 1.
ProDom; PD002078; Fib_collagen_C; 1.
SMART; SM00038; COLFI; 1.
SMART; SM00282; LamG; 1.
SMART; SM00210; TSPN; 1.
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Glycoprotein; Collagen; Signal; Alternative splicing;
Disease mutation.
SIGNAL 1 35
PROPEP 36 511
CHAIN 512 1561
PROPEP 1562 1804
DOMAIN 37 417
DOMAIN 418 506
DOMAIN 507 509
DOMAIN 510 527
DOMAIN 528 1540
DOMAIN 1541 1561
CARBOHYD 1638 1638
SITE 610 610
SITE 1450 1450
VARSPIC 329 413
VARIANT 189 195
VARIANT 196 1804
SEQUENCE 1804 AA; 180963 MW; FE2DB9DED1E4219A CRC64;
POTENTIAL.
AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
COLLAGEN ALPHA 1(XI) CHAIN.
CARBOXYL-TERMINAL PROPEPTIDE.
NONHELICAL REGION.
TRIPLE-HELICAL REGION (INTERRUPTED).
SHORT NONHELICAL SEGMENT.
TELOPEPTIDE.
TRIPLE-HELICAL REGION.
NONHELICAL REGION (C-TERMINAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
CROSSLINKING.
CROSSLINKING.
MISSING (IN SHORT ISOFORM).
LDSERS -> SIEVRDR (IN CHONDRODYSPLASIA).
MISSING (IN CHONDRODYSPLASIA).

```

Alignment Scores:

QY	2158	-	GCGACCCCTAGAGAGATCGGGGAGA
Db	628	pGlyMetArgGlyGluaspGlyGluileGlyProArgGlyLeuProGlyGluAlaGlyPr	648
QY	2135	GAGG-	--AGAGGGACGCCACAGCCCCAGCTGTGCAGCTACGACACCTC
Db	648	oArGglyLeuLuglyProArgGlyThrProGlyProProGly-	662
QY	2090	AGCAGCACAGGTTGGCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAACTTGGCGGCC	2031
Db	663	-----GlnProGlylleGlyGlyleAsp-	670
QY	2030	AGCCCGCAGCCCATGGGCTAACAGGACGGGGAGCTGGGACC--AGTCAGCAGG	1974
Db	671	-----GlyProGlnGlyPro-----LysGlyAsnMetGlyProGlnGlyGluProGl	686
QY	1973	CCCTCCACCCCAATGTCTGGAAGTTTTCTACGCTGAGTAGTTTGGCCAAGTCCTCTGT	1914
Db	686	yProPro-	688
QY	1913	CAAATACTACCTGTGTAGCAAACTAATGCGACCCAGAC---CCAGGCTCGGGCNAC	1858
Db	689	-----GlyGlnGlnGlyAsnProGlySerGlnGlyLe	699
QY	1857	ACCATATAGGCAGTACAGACTGCCTGAGCTGGACAATGGAGCCATAAACAGGATGGG	1798
Db	699	uProGlyProGlnGlyProIleGlyProProGlyGlyLysGlyPro-----GlnGl	716
QY	1797	GCCACCTGGGACAGCAAGGCCTATCCAGGATGCGAGGTCAGGCAGATGCCCGG	1738
Db	716	yLysProGlyLeuAlagly-----LeuProGlyAlaaspGlyProGlyHIsProGl	734
QY	1737	CCCGAA-----CCACCCTGGCCTCGGTGG-----	1713
Db	734	yLyseluglyGlnSerGlyGluLysGlyAlaLeuGlyProProGlyProGlnGlyProIl	754
QY	1712	-GCTCACCCACCACACA-:	1696
Db	754	eGlyTyrrProGlyProArgGlyValLysGlyAlaaspGlyValargGlyLeuLysGlyse	774
QY	1696	-----	1696
Db	774	rLysGlyGluLysGlyGluaspGlyPheProGlyPheLysGlyAspMetGlyLeuLysGl	794
QY	1695	----CGTACGGAGACATCACAGGCA-----GAGGCCCGCA	1664
Db	794	yAspArgGlyGluValGlyGlnValgylProArgGlyGluaspGlyProGluGlyProLy	814
QY	1663	GAGCGCGGTGGAGTGCGGAGCAGGCCACTGCCTCCAGCACCACGTCTCCATTAGGAA	1604
Db	814	sGlyArgAlaGlyProThrGlyAspProGlyProserGlyGlnAlaGlyLysGlyLy	834
QY	1603	G---GGAGCTCCAGCTTAGGG-----CCTGGCAGGAGCTGGTCATCAGGCTGCTCTC	1555
Db	834	sLeuGlyValProGlyLeuProGlyTyrrProGlyArgGln-----	847
QY	1552	ACTGTAGCACCTCCAGTGTCCCTCGGTATTTGGCAGGAAACCTGTCTTCCCCTGGT	1499
Db	848	-----GlyProLysGlySer-ThrGlyPheProGlyPheProGlyA	861
QY	1492	GTAGAGGAGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAAGTGAACCCCGGTGAG	1433
Db	861	IaasnGlyGluLysGlyAlaArg--GlyleAlaGly-----	872
QY	1432	GGCGGTGAAGCTGTACACACGCCNCACCTGTGGCAGCAGCATGTGGCACCCGACCCAC	1377
Db	873	-----LysProGlyProArgGlyGlnArgGly-----	881
QY	1372	AGGGAAGCTGCCACACTGCCCAATAGACTGCTCGAGTCCGGAATCGCTGCACCAAGCGG	1311
Db	882	-----	ProThrcG 884
QY	1312	GTCCA-----TGACCGAGAGAGACACCGAGGAGATGGCGCACTG	1274

```

Db      884  lyProArgGlySerArgGlyAlaArgGlyProThrGlyLysProGlyProLysGlyThrs 904
QY      1273  CAGGA-----ACAGCCCGAGGCTGCCCGGAGCGCTTATCATAGTGTCTCGGGC 1220
Db      904  erGlyGlyAspGlyProGlyProGlyProGlyGluArg-----Glyp 918
QY      1219  CTGGTGGCGGCTAGCTCTGGGCACGCCCTGGTACAGCCCTGCCCGCACGAATCGGT 1160
Db      918  roGlnGlyPro-----GlnGlyProValGlyPheProGlyProLysGlyProp 934
QY      1159  GTAACACAGCTGAAGTTCATGAGTGCCTCCAGCTGCACAGCTGCGGAGGAGCGGCCAG 1100
Db      934  roGlyProAlaGlyLysAspGlyLeuProGlyHis----- 945
QY      1099  GGCAGGGTGGCGGCATCGGCAGCAGCTGTGTCAGCCGGGAGGAGCGGCCCGCCAG 1040
Db      946  -----ProGlyGlnArgGlyGluThrGlyPheGlnG 956
QY      1039  GTTCCGGAAGCAAGCGGCCCGCG-----ATGCACAGCAGTGGGCGCACAGAGGG 986
Db      956  lyLysThrGlyProGlyProGlyGlyValValGlyProGlnGlyProThr-----G 974
QY      985  GCGCACACCCCTCTCTGGTGGTGGTGGGCGGCGGCTGCTCTCAGCCAGCAGCAG 926
Db      974  lyGluThrGlyProIleGlyGluArgGlyHisProGly-----ProGlyProProGly 993
QY      925  TGTGCTGCTAGCAGGTGAGGAAGATGAGGTGAGCAGGCGCAAGAGCAGCTCTCT----- 870
Db      993  luGlnGlyLeuPro-----GlyAlaAlaGlyLysGluGlyAlaLysGlyAspProGlyp 1011
QY      869  -----CCTGGGTGGCCAGGT---AGGG 851
Db      1011  roGlnGlyLysSerGlyLysAspGlyProAlaGlyIleArgGlyPheProGlyGluArg 1031
QY      850  GG---CCAGGGCAGCTGTGTCCATGTCAGCAGGAGGAGGTAGCCAGGCGGCCCGCC 794
Db      1031  lyLeuProGlyAlaGlnGlyAlaProGlyLeuLysGlyGlyGluGlyProGlnGlyPro 1051
QY      793  AAGACTGATCATGAGGATAGACAGAGTGGCTGGGCGGACAGTGGTCCGGT----- 741
Db      1051  lngly-----ProValGlySerProGlyGluArgGlySerAlaGlyAla 1066
QY      740  -----CCCGGAAGAGTCCAGAGCAGGCGGCTCCCA-----GTGG 707
Db      1066  laglyProIleGlyLeuProGlyArgProGlyProGlnGlyProGlyProGlyProAla 1086
QY      706  AGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCGCCAGC----- 666
Db      1086  luLysGlyAlaProGlyGluLysGlyProGlnGlyProAlaGlyArgAspGlyValGln 1106
QY      665  -----CCAGGATGACAGTGCAGCT-----CCAGGG----- 639
Db      1106  lyProValGlyLeuProGlyProAlaGlyProAlaGlySerProGlyGluAspGlyAsp 1126
QY      638  -----GCCTGGATCCGGGCACAGCCCTGCTAGCCAGCCCG----- 600
Db      1126  ysGlyGluIleGlyGluProGlyGlnLysGlySerLysGlyAspLysGlyGluAsnGlyp 1146
QY      599  -----CCCTTGGGATGAGAAGAGGCTCAGCAGGAGTCCCAAGCAGCAGTCCCGAT 548
Db      1146  roProGlyProProGly-----LeuGlnGlyProValGlyAlaProGlyIle----- 1161
QY      547  GAAGGGCGGGCGGCGCATAGCGTCCAGCCAGTGGTCTAGCTGGCTGAGCCCTAGGA----- 492
Db      1162  -----AlaGlyGlyAspGlyGluProGlyProArgGlyGlnGlnGlyMetPheGlyGln 1180
QY      491  -----GGGGCACACAGA-----CCAGGCCCA-----GCAGTGGACCAAT 458
Db      1180  ysGlyAspGluGlyAlaArgGlyPheProGlyLeuProGlyProIleGlyLeuGlnGlyL 1200
QY      457  GCCCAGCAGCATGGTCATCACTCT-----CCTC 428

```

```

1200 euProGlyProProGlyGluLysGlyGluAsnGlyAspValGlyProMetGlyProProG 1220
427 TACCCCTTCCTCCAGCAGCAGAGCGGCACATAGTGTATGCTCGCGCCAAACACACTC 368
1220 lyProProGlyProArgGlyProGlnGlyProAsnGlyAlaAspGlyProGlnGlyPro 1240
367 CAGGCCAA-----AGGTTAGCAGGTTGACCAAGAGAGCTGGGCTTTCCGGTG 320
1240 roGlySerIleGlySerValGlyValValGly-AspLysGlyGlu----- 1254
319 CCGCAGCAGCGCGCTCACCCACAGCCTCTGACCATAGTGGGCCAGCGGG----- 269
1255 ProGlyGluAlaGlyAsnProGlyProProGlyGluAlaGlySerGlyLeuLysGly 1274
268 ---TAGGCTCAGGGGCGGTTCCAGGCACCTCCAGAACTGCTCTCTCGGCTCTGCTCCA 212
1275 GluArgGlyGluLysGlyGluAlaGlyProPro-----GlyAlaAlaGly 1289
211 GAAGCTCGCGGCTCTCTCTCTCTGCTGCGCGCACTGCTAGGAATCAGCAGGCGCCAT 152
1290 ProAlaGlyLysGlyProProGlyAspAspGlyProLysGlyAsnProGly----- 1307
151 TTCTGCCAGCGCTTTGGT-----GCCGTCCTCAGCTTCTCAGCCCATGCTCAAC 104
1308 -----ProValGlyPheProGlyAspProGlyProProGlyGluPro-----Gly 1322
103 ACCTGCTGCTCTGCGGCACCTCAGTGGGACACGCTCATCATCAGTCTGCTGCGCA-- 46
1323 -ProAlaGlyGlnAspGlyValGlyGlyAspLysGlyGluAspGlyAspProGlyGlnPr 1342
45 -----GCCGCGCGGCTGTC 32
1342 oGlyProProGlyProSerGlyGluAlaGlyProProGlyProProGlyLysArgGlyPr 1362
31 ACCCGAGCGCCAGC 19
1362 oProGlyAlaSer 1366

```

RESULT 44
CALA_MOUSE
ID CALA_MOUSE STANDARD; PRT; 680 AA.
AC Q05306;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Collagen alpha 1(X) chain precursor.
GN COL10A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=93143676; PubMed=8424763;
RA Elima K., Berola I., Rosati R., Metsaranta M., Garofalo S., Perala M.,
de Crombrughe B., Vuorio E.;
RT "The mouse collagen X gene: complete nucleotide sequence, exon
structure and expression pattern.";
RL Biochem. J. 289:247-253(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=Liver;
RX MEDLINE=93238750; PubMed=8477738;
RA Kong R.Y.C., Kwan K.M., Lau E.T., Thomas J.T., Boot-Handford R.P.,
Grant M.E., Cheah K.S.E.;
RT "Intron-exon structure, alternative use of promoter and expression of
the mouse collagen X gene, Col10a-1.";
RL Eur. J. Biochem. 213:99-111(1993).
RN [3]
RP SEQUENCE OF 51-680 FROM N.A.
RC STRAIN=DBA/2J;
RX MEDLINE=92267014; PubMed=1587271;

Db 489 -GlyAlaGlnThrLeuPro-----GlySerAspSerValSerThrTrp----- 502
QY 2631 CTTCTACTTTGTTAAATAAATAGTTAAATATTAATGCCCTGTCTCTGTGATGGCAA 2572
Db 503 -----AspGlySe 505
QY 2571 CAGAAGACCAACAGAGCCACATCCTGATAAAAGG---TAAGAGGGGGTGGATCAGCAAA 2515
Db 505 rValArgThr-----ProGlyGlyArgValLysGluGlyLeuLysGlyGly 521
QY 2514 AAGACAGTCTGTGGGCTCAGGGGA-----CCTGGTCTTGTGT 2476
Db 521 nLysGluProGlyValProGlyProGlyArgAlaGlyProProGlySerProCy 541
QY 2475 GTTGCC----- 2469
Db 541 sLeuProGlyProProGlyLeuProCysProValSerProLeuGlyProAlaGlyProAl 561
QY 2468 -CTCAGGACTCTTCCCTTCAAAATAGTCATATGTTCAAATCCCATGGAGGAGTGTTC 2410
Db 561 aLeuGlnThrValPro-----GlyProGlnGly 570
QY 2409 TCTCAAACTCCCATGCAAGACTACATTAAACGAAGCTCAGGTTAAGGGCTTAGAG 2350
Db 570 yProProGlyProProGlyArg-----AspGlyThrProGly-Arg 584
QY 2349 ATGGGAACAGGTGACTGAGTTTATTACGCTCCCAAAACCCCTTCTAGTGTGTCTC 2290
Db 584 spGlyGluProGlyAsp----- 589
QY 2289 ACTAGGAGGTAGCTGTAAACCTGAGCCTGGGTAAATCCACTGCAGAGTCCCGCAT 2230
Db 590 -----ProGlyGluAspGlyLysP 596
QY 2229 CCAGTGCATGGAGCCTTCTGGCCTCCTGTATAGTCCAGACTCAAAACCCCTTGGAG 2170
Db 596 roGlyAspThrGlyProGlnGlyPhePro----- 605
QY 2169 GCCTCCAGTCAGGACCTTAGAGCTGGGAGAGAGG----- 2132
Db 605 -----GlyThrProGlyAspValGlyProLysGlyAspLysGlyAspProGlyV 622
QY 2131 -----AGAGGAGCCTCCAGCCTCCAGCCTCCAGCTGTGAGCTAC 2098
Db 622 alGlyGluArgGlyProProGlyProGlnGlyProProGlyProGlyProSerPheA 642
QY 2097 GCACCTCAGCAGCAGGCTGGCAGCAGAGCCACATTACTTTGGCAGCAACAGAACT 2038
Db 642 rGHisAspLysLeuThrPheIleAspMetGluGlySerGlyGlyGly-----AspL 660
QY 2037 GCGGCCAGCCGCGAGCCCATGGGCTAACAGAGCGGGAGCTGGAGCCAGTGAGG 1978
Db 660 euGluAlaLeuArgGlyProArgGlyPhe-----ProGlyProp 673
QY 1977 CAGGCCCTCCA-----CCCAATGTCTGGAAGTTTCTACCTGAGTATTTGGCCAAG 1924
Db 673 roGlyProProGlyValProGlyLeuProGlyGluProGlyArg-----PheGly 689
QY 1923 TCGCTCTGTGCAATACTACTGTGTAGCAAAAGTAAATGGCGA-----CCAGACCCA 1872
Db 690 -----ValAsnSerSerAspValProGlyProA 699
QY 1871 -GGCCTCGCGCACACCATATAGGAGTGCACAGACTGGCTGAGCTGGACATGGAGCC 1814
Db 699 laGlyLeuProGlyValProGlyArgGluGlyProProGlyPheProGlyLeu----- 716
QY 1813 CATAACAGGATGGGCGACCTGGGACAGCAGGAGGACCTATCCA-----G 1766
Db 717 -----ProGlyProProGlyProProGlyArgGluGlyProProGlyArgThrG 733
QY 1765 GATGGCAGGTCCA-----GGCAGATGCCCGCGCGCAACCACTGGCCTCGCTGGGCTC 1709
Db 733 lyGlnLysGlySerLeuGly-GluAlaGlyAlaProGlyHisLysGly-SerLysGlyAl 752

QY 1708 A-----CCACCACACACAGTGGAGACATCACAGCAGAGCCCGCAGAGCGCGGG 1655
Db 752 aproGlyProAlaGlyAlaArgGlyGluSerGlyLeuAlaGlyAlaProGlyProAlaGly 772
QY 1654 TGGAGTGGGAGCAGGCCACTGCTCCAGCACCACCGTGTCCATTAGGGAAGGAGCTCC 1595
Db 772 yProProGly-----ProProGlyProProGlyProProGlyProGlyLeuPr 788
QY 1594 AGGCTTAGGGCTGCGCAGGAAGTGGTCATCAGCTGTCCCTCAGCTAGTACCTCCAGT 1535
Db 788 o-Ala--GlyPheAspMetGluGlySerGlyGlyPro----- 800
QY 1534 GTCCTCGGTATTGGGCA-----GGAACACCTGCTT 1502
Db 801 -----PheTrpSerThrAlaArgSerAlaAspGlyProGlnGlyProProGlyL 817
QY 1501 CTCCTGGTGGAGAGGAGGCCAGTGTGTAGGCGAGGATCTGCAGGCTGAGAAGGTGAA 1442
Db 817 euProGlyLeuLysGlyAspPro-----GlyValProGlyLeu----- 829
QY 1441 CCCGTGAGGCGCTGAAGCTGTACACGGCCACACTGTGGCAGCAGCATGTGGCACC 1382
Db 829 ----- 829
QY 1381 GGCAGCCACAGGGAAG-----CTGCCACACTGGCCAAATAGAC 1343
Db 830 -----ProGlyAlaLysGlyGluValGlyAlaAspGlyIleProGlyPheProGlyLeuP 848
QY 1342 TGCTCGAGTCCGCAATCGTCACAGCCGCTCATCACCAGCAGAGA----- 1296
Db 848 roGlyArgGluGlyIleAlaGlyProGlnGlyProLysGlyAspArgGlySerArgGlyG 868
QY 1295 -----AGACCAGGAGATGGCGCAGCTGCAGGAA 1268
Db 868 luLysGlyAspProGlyLysAspGlyValGlyGlnProGly-----LeuProG 884
QY 1267 CAGCCCGAGCTGCCATCCGAGCGCTTCAT-----CAT 1232
Db 884 lyProProGlyProProGlyProValValThrValSerGluGlnAspGlySerValLeuS 904
QY 1231 GTGTCTCGGCTCGGTGCGCGCTCAGCTCTGGCAGCCCTGTGTACAGCC-----C 1178
Db 904 erValProGlyProGlu-----GlyArgProGlyPheAlaGlyPheP 918
QY 1177 CTGCCCCACAAATCCGTGTAAACAGCGTGAAGGTCTAGTGCATCCAGCTGCACAG 1118
Db 918 roGlyProAlaGlyPro----- 923
QY 1117 CTCAGCCACGAAGAGCGCGGCGAGGTGGGGCATCGGCAGCAGCTGGTGCAGCCG 1058
Db 924 -----LysGlyAsnLeuGlySerLysGlyGluArgGlySerProGlyProLysG 940
QY 1057 GGAAGCAGGCGCCAGGTTC-----GGAAGCCAAAGCGCGCCGATGAGCAGCAGTG 1001
Db 940 lyGluLysGlyGluProGlySerIlePheSerProAspGlyGlyAlaLeu-----G 957
QY 1000 GSGCAGACAGGAGGCGCCAGCAGCCCTTCTGCTGCTCGTGGGCGCCAGCGCTG----- 945
Db 957 lyProAlaGlnLysGly-----AlaLysGlyGluProGlyPheArgG 971
QY 944 --CCTCCTCAGCCACAGCAGTGTGCTGTCTAGCAGGTGAGAGCATCAGGTGAGCAG 887
Db 971 lyProProGlyProTy-----GlyArgProGlyTyrLysG 983
QY 886 GCCAAGAGGACTCTCTCTGGTGGCCAGGTAGGCGCCAGGCGCAGTGTGCTCCAGTC 827
Db 983 lyGluIleGlyPhe-----ProGlyArgProGlyArg-----ProGlyM 996
QY 826 AATGGCAGCAGGAGTACCCAGCAGCCCGCCAGCAGCTATCA-----TGAA 779
Db 996 etAsnGlyLeuLysGlyGluLysGlyGluProGlyAspAlaSerLeuGlyPheGlyMeta 1016

Search completed: June 9, 2003, 22:29:29
Job time : 320 secs

DB 2221 ATGCACATGGAATGCGGCACTCTGCAGGTGATATCCAGGCTCAGGGTTACAGTAGC 2280
QY
DB 2281 CTCCTAGCTTGCAGACACACTAGAGAGGGTTTGGGAGCTGAATAAATCACTACCTG 2340
DB 2281 CTCCTAGCTTGCAGACACACTAGAGAGGGTTTGGGAGCTGAATAAATCACTACCTG 2340
QY 2341 GTTTCCTCATCTTAAGCCCTTAACCTGCAGCTTCGTTAAATGCTAGCTCTGCATGGAG 2400
DB 2341 GTTTCCTCATCTTAAGCCCTTAACCTGCAGCTTCGTTAAATGCTAGCTCTGCATGGAG 2400
QY 2401 TTTCTAGGATGAACACTCTCTCATGGGATTTGAACATATGACTTATTTGTAGGGGAAGA 2460
DB 2401 TTTCTAGGATGAACACTCTCTCATGGGATTTGAACATATGACTTATTTGTAGGGGAAGA 2460
QY 2461 GTCTGAGGGGCAACACAAAGAACAGGTCCTTCAGCCACAGCAGCTGTCTTTTGTCT 2520
DB 2461 GTCTGAGGGGCAACACAAAGAACAGGTCCTTCAGCCACAGCAGCTGTCTTTTGTCT 2520
QY 2521 GATCCACCCCTCTTACCTTTATCAGGATGTGGCTGTGTGCTCTCTGTTCATCA 2580
DB 2521 GATCCACCCCTCTTACCTTTATCAGGATGTGGCTGTGTGCTCTCTGTTCATCA 2580
QY 2581 CAGAGACACAGCATTAATATTTAACTTATTTAACTTATTTAACTTATTTAACTTATTTAACT 2640
DB 2581 CAGAGACACAGCATTAATATTTAACTTATTTAACTTATTTAACTTATTTAACTTATTTAACT 2640
QY 2641 TGTAGCTTTTCTGTGCTGCTTAATATTTGGGTAGGTGGGATGCTTCTCTCTGGGT 2700
DB 2641 TGTAGCTTTTCTGTGCTGCTTAATATTTGGGTAGGTGGGATGCTTCTCTCTGGGT 2700
QY 2701 GGTCCCTCAGATAGTGTCTATTTGGGTGATTCATTTGCCAGATCTTCTCTCTGGGT 2760
DB 2701 GGTCCCTCAGATAGTGTCTATTTGGGTGATTCATTTGCCAGATCTTCTCTCTGGGT 2760
QY 2761 CTGGCCCTCCCAAAATGCTTAACCCAGGACCTTGGAAATCTACTCATCCCAAAATGATAAT 2820
DB 2761 CTGGCCCTCCCAAAATGCTTAACCCAGGACCTTGGAAATCTACTCATCCCAAAATGATAAT 2820
QY 2821 TCCAAATGCTTTACCCAGGTTAGGTTGTTGAAGGAAGTGAAGGTTGGGCTTCAGGT 2880
DB 2821 TCCAAATGCTTTACCCAGGTTAGGTTGTTGAAGGAAGTGAAGGTTGGGCTTCAGGT 2880
QY 2881 CTCACGGCTTCCCTTAACCCCTCTTCTCTGTGGCCAGGCTGTTTCCCTCCCTTCCA 2940
DB 2881 CTCACGGCTTCCCTTAACCCCTCTTCTCTGTGGCCAGGCTGTTTCCCTCCCTTCCA 2940
QY 2941 CTCCTCTACTCTCTTAGGACTGGGCTGATGAAGGACCTGCCCAAAATTTCCCTACC 3000
DB 2941 CTCCTCTACTCTCTTAGGACTGGGCTGATGAAGGACCTGCCCAAAATTTCCCTACC 3000
QY 3001 CCCAACTTCCCTTACCCCAACTTTCCCAAGGCTTCCCAAGGCTTCCCAAGGCTTCCCAAGGCT 3060
DB 3001 CCCAACTTCCCTTACCCCAACTTTCCCAAGGCTTCCCAAGGCTTCCCAAGGCTTCCCAAGGCT 3060
QY 3061 GCAGACAGAGACAAAGTGCAGGTTTCCCAAGGCTTCCCAAGGCTTCCCAAGGCTTCCCAAGGCT 3120
DB 3061 GCAGACAGAGACAAAGTGCAGGTTTCCCAAGGCTTCCCAAGGCTTCCCAAGGCTTCCCAAGGCT 3120
QY 3121 ATATCTGTGCTTGGGGAATCTCACACAGAACTCAGGAGCAACCCCTGCTTGAGCTAAGG 3180
DB 3121 ATATCTGTGCTTGGGGAATCTCACACAGAACTCAGGAGCAACCCCTGCTTGAGCTAAGG 3180
QY 3181 GAGTCTTATCTCTCAGGGGGGTTTAAAGTCCCTTTGCAATATATGCTTATTTATTT 3240
DB 3181 GAGTCTTATCTCTCAGGGGGGTTTAAAGTCCCTTTGCAATATATGCTTATTTATTT 3240
QY 3241 TAGGGGGGTGAATATTTTATCTTAAGTGAGCAATCAGAGTATATGTTTATGTTGACA 3300
DB 3241 TAGGGGGGTGAATATTTTATCTTAAGTGAGCAATCAGAGTATATGTTTATGTTGACA 3300
QY 3301 AAATTAAGGCTTCTTATATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3360

DB 3301 AAATTAAGGCTTCTTATATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 3360
QY 3361 AA 3410
DB 3361 AA 3410

RESULT 2

US-09-030-607-110
; Sequence 110, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiaochun
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-09-030-607-110

Query Match 100.0%; Score 3409.6; DB 4; Length 3410;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAACACAGCTCTCAGCGCTGCTCCGGGTGACAGCCGCGCCCTCGCCAGGATCTCA 60
DB 1 GGGAACACAGCTCTCAGCGCTGCTCCGGGTGACAGCCGCGCCCTCGCCAGGATCTCA 60
QY 61 GTGATGAGAGCTGTCCCTACTGAGTGTGCCCCACAGCAGCAGGTGTTGAGCATGGCTGAG 120
DB 61 GTGATGAGAGCTGTCCCTACTGAGTGTGCCCCACAGCAGCAGGTGTTGAGCATGGCTGAG 120
QY 121 AAGCTGAGCGCGGACCAAGGCTGGCAAAATGGCGCTGGCTGATTCCTAGGAGTT 180
DB 121 AAGCTGAGCGCGGACCAAGGCTGGCAAAATGGCGCTGGCTGATTCCTAGGAGTT 180
QY 181 GGCGGACGCAAGGAGGAGCGCGGCTTCTGGAGCAGAGCCGAGCAGTTCCTG 240
DB 181 GGCGGACGCAAGGAGGAGCGCGGCTTCTGGAGCAGAGCCGAGCAGTTCCTG 240
QY 241 GAGTGCCTCAACGGGCCCCCTGAGCCCTACCCGCCCTGGCCCCACTATGGTCCAGAGCTGTG 300
DB 241 GAGTGCCTCAACGGGCCCCCTGAGCCCTACCCGCCCTGGCCCCACTATGGTCCAGAGCTGTG 300

QY 661 CTTGGGCGTGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCT 720
DB |||||||
QY 661 CTTGGGCGTGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCT 720
DB |||||||
QY 721 GCTCTCTGACCTCTTCCGGGACCGGACACTGTGCGCAGGCGTACTGTCTATGCTT 780
DB |||||||
QY 721 GCTCTCTGACCTCTTCCGGGACCGGACACTGTGCGCAGGCGTACTGTCTATGCTT 780
DB |||||||
QY 781 CATGATCAGTCTTGGGGCTGCCTGGGTACTCTCTGCTGCCATGTGACTGGGACACAG - 840
DB |||||||
QY 781 CATGATCAGTCTTGGGGCTGCCTGGGTACTCTCTGCTGCCATGTGACTGGGACACAG 840
DB |||||||
QY 841 TGCCTTGGGCGCTTACTTGGGACCCAGGAGGTGCTCTTGGCTGCTGCTCAGCCCTCAT 900
DB |||||||
QY 841 TGCCTTGGGCGCTTACTTGGGACCCAGGAGGTGCTCTTGGCTGCTGCTCAGCCCTCAT 900
DB |||||||
QY 901 CTTCTCTCAGCTCGGTAGCAGCCACACTGCTGTGGCTGAGGAGGAGCGCTGGGCGCCAC 960
DB |||||||
QY 901 CTTCTCTCAGCTCGGTAGCAGCCACACTGCTGTGGCTGAGGAGGAGCGCTGGGCGCCAC 960
DB |||||||
QY 961 CGAGCCAGCAGAGGCGTGTGGGCGGCTCTTGTGCGCCCACTGCTGCTCATGCCGGC 1020
DB |||||||
QY 961 CGAGCCAGCAGAGGCGTGTGGGCGGCTCTTGTGCGCCCACTGCTGCTCATGCCGGC 1020
DB |||||||
QY 1021 CGGCTTGGCTTTCGGGAACCTGGGCGGCTGCTTCCCGGCTGCAACAGCTGTGCTGCG 1080
DB |||||||
QY 1021 CGGCTTGGCTTTCGGGAACCTGGGCGGCTGCTTCCCGGCTGCAACAGCTGTGCTGCG 1080
DB |||||||
QY 1081 CATGCCCGCAGCCCTGCGCGGCTCTTGTGGCTGAGCTGTGCACTGGATGGCACTCAT 1140
DB |||||||
QY 1081 CATGCCCGCAGCCCTGCGCGGCTCTTGTGGCTGAGCTGTGCACTGGATGGCACTCAT 1140
DB |||||||
QY 1141 GACCTTCAGCTGTTTACACGATTCGTGGCGAGGCGCTGACAGGCGTCCCGCAG 1200
DB |||||||
QY 1141 GACCTTCAGCTGTTTACACGATTCGTGGCGAGGCGCTGACAGGCGTCCCGCAG 1200
DB |||||||
QY 1201 AGCTGACCGGCGCAGCGGCGGAGACACTATGATGAAGCGCTTCGATGGGAGCGCT 1260
DB |||||||
QY 1201 AGCTGACCGGCGCAGCGGCGGAGACACTATGATGAAGCGCTTCGATGGGAGCGCT 1260
DB |||||||
QY 1261 GGGGCTGTTCTGCACTGGCCATCTCCCTGCTCTCTCTGCTGCTATGACCGCGCTGT 1320
DB |||||||
QY 1261 GGGGCTGTTCTGCACTGGCCATCTCCCTGCTCTCTCTGCTGCTATGACCGCGCTGT 1320
DB |||||||
QY 1321 GCAGCATTCGGCACTCGACAGTCTATTGGCCAGTGTGGAGCTTTCCTCTGCTGCTGC 1380
DB |||||||
QY 1321 GCAGCATTCGGCACTCGACAGTCTATTGGCCAGTGTGGAGCTTTCCTCTGCTGCTGC 1380
DB |||||||
QY 1381 CGGTGCCACATGCTGTCCACAGTGTGGCCGCTGTGACAGCTTTCAGCGCGCTCACCGG 1440
DB |||||||
QY 1381 CGGTGCCACATGCTGTCCACAGTGTGGCCGCTGTGACAGCTTTCAGCGCGCTCACCGG 1440
DB |||||||
QY 1441 GTTCACCTTCTGAGCCCTGAGATCTGCTTACACACTGGCTTCCCTTACACCGGGA 1500
DB |||||||
QY 1441 GTTCACCTTCTGAGCCCTGAGATCTGCTTACACACTGGCTTCCCTTACACCGGGA 1500
DB |||||||
QY 1501 GAAGCAGGTGTTCTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAAGACAG 1560
DB |||||||
QY 1501 GAAGCAGGTGTTCTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAAGACAG 1560
DB |||||||
QY 1561 CCTGATGACAGCTTCTGCGAGCGCTTACGCTTGGAGTCCCTTCCCTTAATGGACAGCT 1620
DB |||||||
QY 1561 CCTGATGACAGCTTCTGCGAGCGCTTACGCTTGGAGTCCCTTCCCTTAATGGACAGCT 1620
DB |||||||
QY 1621 GGGTGTGGAGGAGTGGCTGCTTCCACCTTCCACCGCGCTCTGCGGGGCTCTGCTG 1680
DB |||||||
QY 1621 GGGTGTGGAGGAGTGGCTGCTTCCACCTTCCACCGCGCTCTGCGGGGCTCTGCTG 1680
DB |||||||
QY 1681 TGATGCTCCGCTAGCTGTGTGGTGGGTGAGCCACCGAGGCGAGGCTGTTCCGGGCGG 1740
DB |||||||
QY 1681 TGATGCTCCGCTAGCTGTGTGGTGGGTGAGCCACCGAGGCGAGGCTGTTCCGGGCGG 1740
DB |||||||

QY 1741 GGCATCTGCTGGAGCTCGCCATCTGATAGTGCTTCTGCTGTCCAGGTGGCCCC 1800
DB |||||||
QY 1741 GGCATCTGCTGGAGCTCGCCATCTGATAGTGCTTCTGCTGTCCAGGTGGCCCC 1800
DB |||||||
QY 1801 ATCCCTGTTTATGGGCTCCATCTGCTCAGCTCAGCAGTCTGCTACCTGATATGCTGTC 1860
DB |||||||
QY 1801 ATCCCTGTTTATGGGCTCCATCTGCTCAGCTCAGCAGTCTGCTACCTGATATGCTGTC 1860
DB |||||||
QY 1861 TGCCGAGGCTGGGCTGCTGCTGCCATTTACTTGTGTACACAGTGTAGTATTTGACAAGAG 1920
DB |||||||
QY 1861 TGCCGAGGCTGGGCTGCTGCTGCCATTTACTTGTGTACACAGTGTAGTATTTGACAAGAG 1920
DB |||||||
QY 1921 CGACTTGGCCAAATACTCAGCTAGAAAATCTCCAGACATTTGGGTGAGGCGCTGCTT 1980
DB |||||||
QY 1921 CGACTTGGCCAAATACTCAGCTAGAAAATCTCCAGACATTTGGGTGAGGCGCTGCTT 1980
DB |||||||
QY 1981 CACTGGGTCCAGCTCCCGCTCTGTTAGCCCATGGGCTGCCGGCTGGCGGCGAGT 2040
DB |||||||
QY 1981 CACTGGGTCCAGCTCCCGCTCTGTTAGCCCATGGGCTGCCGGCTGGCGGCGAGT 2040
DB |||||||
QY 2041 TTCTGTTGCTGCCAAAGTAAATGTGGCTCTCTGTGCTGCCACCTGTGCTGAGTGGCTA 2100
DB |||||||
QY 2041 TTCTGTTGCTGCCAAAGTAAATGTGGCTCTCTGTGCTGCCACCTGTGCTGAGTGGCTA 2100
DB |||||||
QY 2101 GCTGCACAGCTGGGGCTGGGGCTCCCTCTCTCTCTCTCCAGCTCTTACAGGCTGCTG 2160
DB |||||||
QY 2101 GCTGCACAGCTGGGGCTGGGGCTCCCTCTCTCTCTCTCCAGCTCTTACAGGCTGCTG 2160
DB |||||||
QY 2161 ACTGGAGGCTTCCAAAGGGGTTTCACTGTGGACTTATACAGGAGGCGCCAGAGGCTCC 2220
DB |||||||
QY 2161 ACTGGAGGCTTCCAAAGGGGTTTCACTGTGGACTTATACAGGAGGCGCCAGAGGCTCC 2220
DB |||||||
QY 2221 ATGCACTTGAATCGGGGACTCTGCACTGGATTTACCCAGGCTCAGGGTTTAAACGTAGC 2280
DB |||||||
QY 2221 ATGCACTTGAATCGGGGACTCTGCACTGGATTTACCCAGGCTCAGGGTTTAAACGTAGC 2280
DB |||||||
QY 2281 CTCTAGTTGAGACACACTCTCCATGGATTTTGGGAGCTGAATAAATCACTCAGTCACTG 2340
DB |||||||
QY 2281 CTCTAGTTGAGACACACTCTCCATGGATTTTGGGAGCTGAATAAATCACTCAGTCACTG 2340
DB |||||||
QY 2341 GTTTCCTCATCTTAAGCCCTTAACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
DB |||||||
QY 2341 GTTTCCTCATCTTAAGCCCTTAACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
DB |||||||
QY 2401 TTTCTAGATGAACACTCTCCATGGATTTTGAACATATGACTTATTTGTAGGGGAAGA 2460
DB |||||||
QY 2401 TTTCTAGATGAACACTCTCCATGGATTTTGAACATATGACTTATTTGTAGGGGAAGA 2460
DB |||||||
QY 2461 GTCTGAGGGGCAACACAAAGAACCCAGGTCCCTCAGCCCAACAGCACTGTCTTTTGTCT 2520
DB |||||||
QY 2461 GTCTGAGGGGCAACACAAAGAACCCAGGTCCCTCAGCCCAACAGCACTGTCTTTTGTCT 2520
DB |||||||
QY 2521 GATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTTGGTCTGCTGCTGCTGCTGCTGCT 2580
DB |||||||
QY 2521 GATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTTGGTCTGCTGCTGCTGCTGCTGCT 2580
DB |||||||
QY 2581 CAGACACAGCATTTAAATATTTAACTTATTTTAAACAAAGTGAAGGGGAATCCAT 2640
DB |||||||
QY 2581 CAGACACAGCATTTAAATATTTAACTTATTTTAAACAAAGTGAAGGGGAATCCAT 2640
DB |||||||
QY 2641 TGCTAGCTTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2700
DB |||||||
QY 2641 TGCTAGCTTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2700
DB |||||||
QY 2701 GGTCCCTGAGATAGCTGGTCAATTTGGGCTGATCTTGGCAGAACTTCTCTCTCTCTCTCTCT 2760
DB |||||||
QY 2701 GGTCCCTGAGATAGCTGGTCAATTTGGGCTGATCTTGGCAGAACTTCTCTCTCTCTCTCTCT 2760
DB |||||||
QY 2761 CTGGCCCCCAAAATGCCTTAACCCAGGACCTTGGAAATTTCTACTCATCCCCAAATGATAT 2820
DB |||||||
QY 2761 CTGGCCCCCAAAATGCCTTAACCCAGGACCTTGGAAATTTCTACTCATCCCCAAATGATAT 2820
DB |||||||
QY 2821 TCCAAATGCTGTTCACCAAGGTTAGGGTGTGAAGGAAGGTAGAGGGTGGGGCTTCAAGT 2880
DB |||||||

Db 2821 TCCAAATGCTGTACCAAGGTTAGGGTGTGAAGAGGTTAGAGGTTGGGCTTCAGGT 2880
QY 2881 CTCACGGCTTCCCTAACACCCCTCTCTCTTGGGCCAGCGTGGTTCCTCCCTCCCA 2940
Db 2881 CTCACGGCTTCCCTAACACCCCTCTCTCTTGGGCCAGCGTGGTTCCTCCCTCCCA 2940
QY 2941 CTCCTCTACTCTCTTAGGACTGGGCTGATGAAGGACCTGCCCAAAATTTCCCTACC 3000
Db 2941 CTCCTCTACTCTCTTAGGACTGGGCTGATGAAGGACCTGCCCAAAATTTCCCTACC 3000
QY 3001 CCCAACTTCCCTTACCCCAACTTCCCAAGCTTCCCAAGCTTCTTGGAGCTACT 3060
Db 3001 CCCAACTTCCCTTACCCCAACTTCCCAAGCTTCTTGGAGCTACT 3060
QY 3061 CGAGGACCAAGGACCAAGGCTTCCCAAGCTTCTTGGAGCTACT 3120
Db 3061 CGAGGACCAAGGACCAAGGCTTCCCAAGCTTCTTGGAGCTACT 3120
QY 3121 ATATCTGTGCTGGGAAATCTCACAGAAACTCAGGAGCAACCCCTGCTGAGCTAAG 3180
Db 3121 ATATCTGTGCTGGGAAATCTCACAGAAACTCAGGAGCAACCCCTGCTGAGCTAAG 3180
QY 3181 GAGGCTTATCTCTCAGGGGGGTTAAGTCCCTTTGCAATAATGTCGTTATTATT 3240
Db 3181 GAGGCTTATCTCTCAGGGGGGTTAAGTCCCTTTGCAATAATGTCGTTATTATT 3240
QY 3241 TAGCGGGGTGAATTTATATCTTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA 3300
Db 3241 TAGCGGGGTGAATTTATATCTTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA 3300
QY 3301 AAATAAGGCTTCTTATATGTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 3360
Db 3301 AAATAAGGCTTCTTATATGTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 3360
QY 3361 AA 3410
Db 3361 AA 3410

RESULT 4

US-09-439-313-110
; Sequence 110, Application US/09439313
; Patent No. 6329505

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 210121.427C9

; CURRENT APPLICATION NUMBER: US/09/439,313

; NUMBER OF SEQ ID NOS: 575

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 110

; LENGTH: 3410

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-439-313-110

Query Match

Best Local Similarity 100.0%; Score 3409.6; DB 4; Length 3410;

Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAACACAGCCTGACACGCTGGCTCGGGTGACAGCCGCGCTCGGGCAGGATCTGA 60
Db 1 GGAACACAGCCTGACACGCTGGCTCGGGTGACAGCCGCGCTCGGGCAGGATCTGA 60
QY 61 GTGATGAGAGCTGTCCTCCCTAGGCTGAGGTCCTCCAGCAGCAGGAGTGTGAGCATGGGCTGAG 120
Db 61 GTGATGAGAGCTGTCCTCCCTAGGCTGAGGTCCTCCAGCAGCAGGAGTGTGAGCATGGGCTGAG 120
QY 121 AAGTGTGACCGGACCAAAAGGCTGGCAGAAATGGGCGCTGCTGATTCCTAGGAGTT 180
Db 121 AAGTGTGACCGGACCAAAAGGCTGGCAGAAATGGGCGCTGCTGATTCCTAGGAGTT 180
QY 181 GGGGCGCAAGAGGAGGAGGCGCAGCTTCTGGAGCAGAGCCGAGCAGGAGGAGGAGTT 240
Db 181 GGGGCGCAAGAGGAGGAGGCGCAGCTTCTGGAGCAGAGCCGAGCAGGAGGAGGAGTT 240
QY 241 GAGTGTGACCGGACCAAAAGGCTGGCAGAAATGGGCGCTGCTGATTCCTAGGAGTT 300
Db 241 GAGTGTGACCGGACCAAAAGGCTGGCAGAAATGGGCGCTGCTGATTCCTAGGAGTT 300
QY 301 GGTGAGCGGCTGCTGCGGACCGGAAAGCCAGCTTCTGCTGCTCAACTGCTAACCTT 360
Db 301 GGTGAGCGGCTGCTGCGGACCGGAAAGCCAGCTTCTGCTGCTCAACTGCTAACCTT 360
QY 361 TGGCCTGGAGGCTGTTTGGCCGAGGATCAGCTATGTCGCGCTCTCTGCTGGAAGT 420
Db 361 TGGCCTGGAGGCTGTTTGGCCGAGGATCAGCTATGTCGCGCTCTCTGCTGGAAGT 420
QY 421 GGGGTGAGAGGAGGAGTTCATGACCATGCTGGGCAATGTCGAGCTGCTGGGCTGCT 480
Db 421 GGGGTGAGAGGAGGAGTTCATGACCATGCTGGGCAATGTCGAGCTGCTGGGCTGCT 480
QY 481 CTGTGTCGCTCTAGGCTCAGCCAGTACCATGCTGGGCAATGTCGAGCTGCTGGGCTGCT 540
Db 481 CTGTGTCGCTCTAGGCTCAGCCAGTACCATGCTGGGCAATGTCGAGCTGCTGGGCTGCT 540
QY 541 GCGCTTCATCTGGGCACTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 600
Db 541 GCGCTTCATCTGGGCACTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 600
QY 601 CGGCTGGCTAGCAGGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 660
Db 601 CGGCTGGCTAGCAGGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 660
QY 661 CCGTGGCTGGGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 720
Db 661 CCGTGGCTGGGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 720
QY 721 GCTCTCTGACCTCTTCCGCGGACCGGACCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 780
Db 721 GCTCTCTGACCTCTTCCGCGGACCGGACCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 780
QY 781 CATGATCAGTCTTGGGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 840
Db 781 CATGATCAGTCTTGGGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 840
QY 841 TGGCCTGGGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 900
Db 841 TGGCCTGGGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 900
QY 901 CTTCTCAGCTGCTGAGCAGCAGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 960
Db 901 CTTCTCAGCTGCTGAGCAGCAGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 960
QY 961 CGAGCAGCAGAGGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 1020
Db 961 CGAGCAGCAGAGGCTGCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 1020
QY 1021 CCGCTTGGCTTTCGGAACCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 1080
Db 1021 CCGCTTGGCTTTCGGAACCTGTCGCGGATCTGTCGCGGCTGCTGTCATCCCAAGGC 1080
QY 1081 CATGCCCGCACCTGCGCGGCTCTTCTGCTGCTGAGCTGTCGAGCTGGATGGCACTCAT 1140


```
|||||
1 GGAACACAGCCTGCACGCGCTGGCTCCGGGTGACAGCGCGCCTCGCCAGGATCTGA 60
61 GTGATGAGACGTGTCCCACTAGGTGAGTCCCAACACAGAGAGGTGTGAGCATGGGCTGAG 120
61 GTGATGAGACGTGTCCCACTAGGTGAGTCCCAACACAGAGAGGTGTGAGCATGGGCTGAG 120
121 AAGCTGGACCGCACAAAGGCTGGCAGAAATGGCGCTGGCTGATTTCTAGGACGTT 180
121 AAGCTGGACCGCACAAAGGCTGGCAGAAATGGCGCTGGCTGATTTCTAGGACGTT 180
181 GCGGCGACGAGAGAGGCGCGCAGCTTCTGGAGCAGAGCCGAGAGCAAGCAAGTCTG 240
181 GCGGCGACGAGAGAGGCGCGCAGCTTCTGGAGCAGAGCCGAGAGCAAGCAAGTCTG 240
241 GAGTGCCTGAACGGCGCCCTGAGCCCTACCCGCTGGCCCACTATGTFCCAGAGGCTGTG 300
241 GAGTGCCTGAACGGCGCCCTGAGCCCTACCCGCTGGCCCACTATGTFCCAGAGGCTGTG 300
301 GGTGAGCGCGCTGTGCGGACCGGAAAGCCAGCTTGTGCTCAACCTGCTAACTT 360
301 GGTGAGCGCGCTGTGCGGACCGGAAAGCCAGCTTGTGCTCAACCTGCTAACTT 360
361 TGGCCTGAGAGTGTCTTGGCGGAGGATCACCTATGTGCCGCTCTGCTGTGGAAGT 420
361 TGGCCTGAGAGTGTCTTGGCGGAGGATCACCTATGTGCCGCTCTGCTGTGGAAGT 420
421 GGGGTAGAGAGAGTTCATGACATGGTCTGGGCAATGGTCAAGTGTGGCGCTGGT 480
421 GGGGTAGAGAGAGTTCATGACATGGTCTGGGCAATGGTCAAGTGTGGCGCTGGT 480
481 CTGTCTCCGCTCTAGGCTCAGCAGTGAACCTAGGCTGTGGTCAACCTGCTAACTT 540
481 CTGTCTCCGCTCTAGGCTCAGCAGTGAACCTAGGCTGTGGTCAACCTGCTAACTT 540
541 GCCCTTATCTGCGGCACTGTCTTGGCATCTCTGAGCTCTTCTCATGCCAAGGCG 600
541 GCCCTTATCTGCGGCACTGTCTTGGCATCTCTGAGCTCTTCTCATGCCAAGGCG 600
601 CGGCTGGCTAGCAGGCTGTGTGCCGGATGCCAGCCCTGGAGCTGGCACTGTCTCAT 660
601 CGGCTGGCTAGCAGGCTGTGTGCCGGATGCCAGCCCTGGAGCTGGCACTGTCTCAT 660
661 CCTGGCGTGGGCTGTGACTTCTGTGCCAGGTGTGCTTCACTCCACTGAGGCGCT 720
661 CCTGGCGTGGGCTGTGACTTCTGTGCCAGGTGTGCTTCACTCCACTGAGGCGCT 720
721 GCTCTCTGACCTTCCGGGACCGGACCACTGTGCCAGGCGCTACTCTCTATGCTT 780
721 GCTCTCTGACCTTCCGGGACCGGACCACTGTGCCAGGCGCTACTCTCTATGCTT 780
781 CATGATCAGTCTTGGGGCTGGCTGGGCTACCTTCCGCGGCTACTCTGTCTATGCCCT 840
781 CATGATCAGTCTTGGGGCTGGCTGGGCTACCTTCCGCGGCTACTCTGTCTATGCCCT 840
841 TGCCCTGGCCCTTACTCTGGGACCGAGGAGTGCCTTCTTGGGCTGTCTCACTCAT 900
841 TGCCCTGGCCCTTACTCTGGGACCGAGGAGTGCCTTCTTGGGCTGTCTCACTCAT 900
901 CTTCCTACCTGCTGAGCAGCACTGTGTGGCTGAGGAGGAGCGCTGGGCGCCAC 960
901 CTTCCTACCTGCTGAGCAGCACTGTGTGGCTGAGGAGGAGCGCTGGGCGCCAC 960
961 CGAGCAGAGAGGCTGTGCGCCCTCTTGTGCGCCCACTGTGTCCATGCGGGC 1020
961 CGAGCAGAGAGGCTGTGCGCCCTCTTGTGCGCCCACTGTGTCCATGCGGGC 1020
1021 CCCTTTGGCTTCCGGAACCTGGGCGCTGTCTTCCCGGCTGACCACTGTGCTGCG 1080
1021 CCCTTTGGCTTCCGGAACCTGGGCGCTGTCTTCCCGGCTGACCACTGTGCTGCG 1080
1081 CATGCCCGGACCTCGCGCGCTCTTCTGTGCTGAGTGTGAGTGGTGGCACTCAT 1140
|||||
```

```

2221 ATGCACTGGAATGCGGGACTCTCGAGTGGATTACCCAGGCTCAGGGTTAACAGCTAGC 2280
2221 ATGCACTGGAATGCGGGACTCTCGAGTGGATTACCCAGGCTCAGGGTTAACAGCTAGC 2280
2281 CTCCTAGTTGAGACACACTGAGAAAGGTTTTCGGAGCTGAATAAACTCAGTCACCTG 2340
2281 CTCCTAGTTGAGACACACTGAGAAAGGTTTTCGGAGCTGAATAAACTCAGTCACCTG 2340
2341 GTTTCOCATCTCTAAGCCCTTAACTCGAGTTCGTTTAACTAGTCTTTCATGGGAG 2400
2341 GTTTCOCATCTCTAAGCCCTTAACTCGAGTTCGTTTAACTAGTCTTTCATGGGAG 2400
2401 TTTCTAGGATGAACACTCTCCATGGGATTTGAACATATGACTTATTTAGGGGAAGA 2460
2401 TTTCTAGGATGAACACTCTCCATGGGATTTGAACATATGACTTATTTAGGGGAAGA 2460
2461 GTCTCAGGGGCAACACAAAGACAGGTCCTCAGCCACAGCACTGTCTTTTGGT 2520
2461 GTCTCAGGGGCAACACAAAGACAGGTCCTCAGCCACAGCACTGTCTTTTGGT 2520
2521 GATCCACCCCTCTTACCTTTATCAGGATGCGCTTGTGCTCTCTCTGTCATCA 2580
2521 GATCCACCCCTCTTACCTTTATCAGGATGCGCTTGTGCTCTCTCTGTCATCA 2580
2581 CAGAGCACAGGCAATTAATAATTTAACTTATTTAACTAGTGAAGGAATCCAT 2640
2581 CAGAGCACAGGCAATTAATAATTTAACTTATTTAACTAGTGAAGGAATCCAT 2640
2641 TGCTAGCTTTCTGTGTGGTGTCTAATATTTGGGTAGGTTGGGGATCCCAACATCA 2700
2641 TGCTAGCTTTCTGTGTGGTGTCTAATATTTGGGTAGGTTGGGGATCCCAACATCA 2700
2701 GGTCCCTGAGATAGCTGTGTCATTGGCTGATCATTTGCCAGATCTTCTCTCTGGGT 2760
2701 GGTCCCTGAGATAGCTGTGTCATTGGCTGATCATTTGCCAGATCTTCTCTCTGGGT 2760
2761 CTGGCCCCCAAAATGCCTAACCCAGGACTTGGAAATTTCTACTCATCCCAATGATAAT 2820
2761 CTGGCCCCCAAAATGCCTAACCCAGGACTTGGAAATTTCTACTCATCCCAATGATAAT 2820
2821 TCCAAATGCTTTACCCAAAGTTAGGTTGTTGAAGGAAGTGAAGGTTGGGCTTCAGCT 2880
2821 TCCAAATGCTTTACCCAAAGTTAGGTTGTTGAAGGAAGTGAAGGTTGGGCTTCAGCT 2880
2881 CTCACGGCTTCCCTAACCCAGCTCTCTCTTTGGCCAGCTGTTGCCCCCACTTCCA 2940
2881 CTCACGGCTTCCCTAACCCAGCTCTCTCTTTGGCCAGCTGTTGCCCCCACTTCCA 2940
2941 CTCCTCTCTCTCTCTAGGCTGATGAAGGCACTGCCCAAAATTTCCCTTACC 3000
2941 CTCCTCTCTCTCTCTAGGCTGATGAAGGCACTGCCCAAAATTTCCCTTACC 3000
3001 CCCAACTTTCCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACC 3060
3001 CCCAACTTTCCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACC 3060
3061 GCAGGACAGAAAGCAAAAGTGGGTTTCCCAAGCTTTGTCATCTCAGCCCTCAGAGT 3120
3061 GCAGGACAGAAAGCAAAAGTGGGTTTCCCAAGCTTTGTCATCTCAGCCCTCAGAGT 3120
3121 ATATCTGTCTTGGGATCTCACACAGAACTCAGGAGCACCCTGCTGAGCTAAGG 3180
3121 ATATCTGTCTTGGGATCTCACACAGAACTCAGGAGCACCCTGCTGAGCTAAGG 3180
3181 GAGGCTTATCTCTCAGGGGGGTTTAAAGTCCCTTTTGAATAATGCTGCTTATTTATT 3240
3181 GAGGCTTATCTCTCAGGGGGGTTTAAAGTCCCTTTTGAATAATGCTGCTTATTTATT 3240
3241 TAGCGGGGTGAATATTTATCTGTAAGTGAAGCAATCAGAGTAAATGTTTATGCTGACA 3300
3241 TAGCGGGGTGAATATTTATCTGTAAGTGAAGCAATCAGAGTAAATGTTTATGCTGACA 3300

Query Match 100.0%; Score 3409.6; DB 4; Length 3410;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAAACAGCTGCACGCGCTGGCTCCGGGTGACAGCGCGCGCTCGGCAGATCTGA 60
Db 1 GGGAAACAGCTGCACGCGCTGGCTCCGGGTGACAGCGCGCGCTCGGCAGATCTGA 60
QY 61 GTGATGACAGCTGTCCTCCACCTGAGTGCCCCACAGCAGCAGGTGTTGAGCATGGCTGAG 120
Db 61 GTGATGACAGCTGTCCTCCACCTGAGTGCCCCACAGCAGCAGGTGTTGAGCATGGCTGAG 120
QY 121 AAGCTGGACCGGCACAAAGGCTGGCAGAAATGGCGCTGGCTGATTCCTAGCAGTT 180
Db 121 AAGCTGGACCGGCACAAAGGCTGGCAGAAATGGCGCTGGCTGATTCCTAGCAGTT 180
QY 181 GCGGCGCAGGAGGAGGAGCGCCGAGCTTCTGGAGCAGCGCAGAGCAGAGCAGTCTG 240
Db 181 GCGGCGCAGGAGGAGGAGCGCCGAGCTTCTGGAGCAGCGCAGAGCAGAGCAGTCTG 240
QY 241 GAGTGCCTGAACGCGCCCTGAGCCCTACCCGCTGGCCCTGAGCCTATGTTCCAGAGGCTGTG 300
Db 241 GAGTGCCTGAACGCGCCCTGAGCCTACCCGCTGGCCCTGAGCCTATGTTCCAGAGGCTGTG 300
QY 301 GGTGAGCCGCTGCTGCGCAGCCGAGGAGCCAGCTTCTGCTGGTCAACCTGCTAACCTT 360
Db 301 GGTGAGCCGCTGCTGCGCAGCCGAGGAGCCAGCTTCTGCTGGTCAACCTGCTAACCTT 360
QY 361 TGGCTTGGAGGTTGTTTGGCCGAGGATCAGCTATGTCGCCCTCTGCTGCTGGAAGT 420
Db 361 TGGCTTGGAGGTTGTTTGGCCGAGGATCAGCTATGTCGCCCTCTGCTGCTGGAAGT 420
QY 421 GGGGGTAGAGGAGAGTTTCACTGACCATGTTGTTGGGCAATGTTCCAGTCTGGGCTGGT 480
Db 421 GGGGGTAGAGGAGAGTTTCACTGACCATGTTGTTGGGCAATGTTCCAGTCTGGGCTGGT 480
QY 481 CTGTGTCCGCTCTAGGCTCAGCCAGTCCAGCTGCGCTGAGCAGCTATGGCGCGCGCG 540
Db 481 CTGTGTCCGCTCTAGGCTCAGCCAGTCCAGCTGCGCTGAGCAGCTATGGCGCGCGCG 540
QY 541 GCCTTTTCATCTGGGCACTGTCTTGGGCAATCTCTGCTGAGCCTCTTTCTCATCCCAAGGC 600
Db 541 GCCTTTTCATCTGGGCACTGTCTTGGGCAATCTCTGCTGAGCCTCTTTCTCATCCCAAGGC 600

RESULT 7
US-09-232-149A-110
; Sequence 110, Application US/09232149A
; Patent NO. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Jennifer Lynn
; APPLICANT: Micham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; FILE REFERENCE: 210121.427C6
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 3410
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-232-149A-110
```


Db 2761 CTGGCCCGCCCAAAATGCGCTAACCCAGGACCTTGGAAATCTACTCATCCCCAAATGATAAT 2820
QY 2821 TCACAAATGCTGTACCCAGGTTAGGGTGTGAAGGAAGGTAGAGGGTGGGGCTTCAGGT 2880
Db 2821 TCACAAATGCTGTACCCAGGTTAGGGTGTGAAGGAAGGTAGAGGGTGGGGCTTCAGGT 2880
QY 2881 CTCAACGGGTTCCCTAACACACCCCTCTCTCTCTGTGGCCAGCGTGTTCCTCCCTTCCA 2940
Db 2881 CTCAACGGGTTCCCTAACACACCCCTCTCTCTCTGTGGCCAGCGTGTTCCTCCCTTCCA 2940
QY 2941 CTCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGACCTGCCCAAAATTTCCCTTACC 3000
Db 2941 CTCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGACCTGCCCAAAATTTCCCTTACC 3000
QY 3001 CCCAACTTTCCCTTACCCCAAACTTTCCCAACAGCTCCCAACCCCTGTGTGGAGCTACT 3060
Db 3001 CCCAACTTTCCCTTACCCCAAACTTTCCCAACAGCTCCCAACCCCTGTGTGGAGCTACT 3060
QY 3061 GCAGGACCAAGACCAAAAGTGGGTTTCCCAAGCTTTGTCCATCTCAGCCGCCAGAGT 3120
Db 3061 GCAGGACCAAGACCAAAAGTGGGTTTCCCAAGCTTTGTCCATCTCAGCCGCCAGAGT 3120
QY 3121 ATATCTGTCTTGGGAATCTCACACAGAACTCAGGAGCACCCCTGCCTGAGCTAAGG 3180
Db 3121 ATATCTGTCTTGGGAATCTCACACAGAACTCAGGAGCACCCCTGCCTGAGCTAAGG 3180
QY 3181 GAGGCTTATCTCTCAGGGGGGTTTAAAGTGGGCTTTGCAATTAATGTCTTATTATT 3240
Db 3181 GAGGCTTATCTCTCAGGGGGGTTTAAAGTGGGCTTTGCAATTAATGTCTTATTATT 3240
QY 3241 TAGCGGGTGAATATTTTACTGTAAAGTGAAGTCAAGATCAGAGTATATGTTTATGTTGACA 3300
Db 3241 TAGCGGGTGAATATTTTACTGTAAAGTGAAGTCAAGATCAGAGTATATGTTTATGTTGACA 3300
QY 3301 AAATTAAGGCTTCTTATATGTTTAAAAAATAAATAAATAAATAAATAAATAAATAA 3360
Db 3301 AAATTAAGGCTTCTTATATGTTTAAAAAATAAATAAATAAATAAATAAATAAATAA 3360
QY 3361 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3410
Db 3361 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3410

RESULT 8

US-09-605-785-704
; Sequence 704, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriek
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605.785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 704
; LENGTH: 4034
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-605-785-704

Query Match 75.8%; Score 2585.4; DB 4; Length 4034;
Best Local Similarity 82.4%; Pred. No. 0;
Matches 3325; Conservative 0; Mismatches 1; Indels 709; Gaps 3;

QY 4 AACAGCGCTGACGCGCTGGCTCGGGTGACACGCGCGCCCTCGCGGAGGATCTCAGTG 63
Db 1 AACAGCGCTGACGCGCTGGCTCGGGTGACACGCGCGCCCTCGCGGAGGATCTCAGTG 60
QY 64 ATGAGAGCTGTCCCACTAGAGTGCCCAACAGACAGAGGTGTGAGCATGGGCTGAGAAG 123
Db 61 ATGAGAGCTGTCCCACTAGAGTGCCCAACAGACAGAGGTGTGAGCATGGGCTGAGAAG 120
QY 124 CTGACCGCGCACCAAAAGGCTGGCAGAAATGGGCGCTGGCTGATTCCTTAGGAGTGGC 183
Db 121 CTGACCGCGCACCAAAAGGCTGGCAGAAATGGGCGCTGGCTGATTCCTTAGGAGTGGC 180
QY 184 GGCAGCAGGAGGAGCGCGCAGCTTCTGAGCAGAGCGAGACGAAGCAGTCTCTGGAG 243
Db 181 GGCAGCAGGAGGAGCGCGCAGCTTCTGAGCAGAGCGAGACGAAGCAGTCTCTGGAG 240
QY 244 TGCCTGAACGCGCCCTGAGCCCTACCCGCTGGCCCACTATGGTCCAGAGGCTGTGGGT 303
Db 241 TGCCTGAACGCGCCCTGAGCCCTACCCGCTGGCCCACTATGGTCCAGAGGCTGTGGGT 300
QY 304 GAGCGCGCTGTGGGCGACCGAAGCCAGCTCTGCTGTCAACCTGCTAACTTTGG 363
Db 301 GAGCGCGCTGTGGGCGACCGAAGCCAGCTCTGCTGTCAACCTGCTAACTTTGG 360
QY 364 CTGAGAGGTGTGTTGGCGCAGGACATCACCTATGTGCGCCCTCTGCTGGAAGTGGG 423
Db 361 CTGAGAGGTGTGTTGGCGCAGGACATCACCTATGTGCGCCCTCTGCTGGAAGTGGG 420
QY 424 GGTAGAGAGAGTTCATGACCATTTGGCATTTGGCATTTGGCATTTGGCATTTGGCAT 483
Db 421 GGTAGAGAGAGTTCATGACCATTTGGCATTTGGCATTTGGCATTTGGCATTTGGCAT 480
QY 484 TGTCCCGCTCTTAGGCTCAGCAGTGACCACTGGCGGTGAGCGCTATGCGCCGCCCGCC 543
Db 481 TGTCCCGCTCTTAGGCTCAGCAGTGACCACTGGCGGTGAGCGCTATGCGCCGCCCGCC 540
QY 544 CTTCATCTGGGCACCTGCTTGGGCATCTGCTGAGCCTCTTTCTCATCCCAAGGCGCG 603
Db 541 CTTCATCTGGGCACCTGCTTGGGCATCTGCTGAGCCTCTTTCTCATCCCAAGGCGCG 600
QY 604 CTGGCTAGCAGGCTGCTGTGCGCGGATCCAGGCGCCCTGGAGCTGGCACTGCTCATCT 663
Db 601 CTGGCTAGCAGGCTGCTGTGCGCGGATCCAGGCGCCCTGGAGCTGGCACTGCTCATCT 660
QY 664 GGGCGTGGGCTGCTGGACTTCTGTGGCAGAGTGTGCTTCACTCCACTGGAGGCGCTGCT 723
Db 661 GGGCGTGGGCTGCTGGACTTCTGTGGCAGAGTGTGCTTCACTCCACTGGAGGCGCTGCT 720
QY 724 CTCTGACCTCTTCGGGAGCCCGGACACTGTGCGCAGGCTACTCTGTCTATGCCCTCAT 783
Db 721 CTCTGACCTCTTCGGGAGCCCGGACACTGTGCGCAGGCTACTCTGTCTATGCCCTCAT 780
QY 784 GATCAGTCTTGGGGCTGCTGCGGTACCTCTGCTGCGCATTTAGTGGACACCAAGTGC 843
Db 781 GATCAGTCTTGGGGCTGCTGCGGTACCTCTGCTGCGCATTTAGTGGACACCAAGTGC 840
QY 844 CTGCGCCCTTACTCTGGGACCCAGAGGAGTGCCTCTTTGGCCTGCTCACCTCATCTT 903
Db 841 CTGCGCCCTTACTCTGGGACCCAGAGGAGTGCCTCTTTGGCCTGCTCACCTCATCTT 900
QY 904 CCTCAGCTGGTAGCAGCCACACTGCTGTGGCTGAGGAGGAGCGCTGGGCGCCACCGA 963
Db 901 CCTCAGCTGGTAGCAGCCACACTGCTGTGGCTGAGGAGGAGCGCTGGGCGCCACCGA 960

964	QY	GCCACAGAAGGGCTGTGGGCCCCCTCCTTTGTCGCCCACTGCTGTCTCCATGCCGGGCCG	1023
961	Db		
961	Db	GCCACAGAAGGGCTGTGGGCCCCCTCCTTTGTCGCCCACTGCTGTCTCCATGCCGGGCCG	1020
1024	QY	CTTGGCTTTCGGGAACCTGGGGCCCTGCTTTCGCCGGCTGCACACAGTGTGCTGCCGCAT	1083
1021	Db		
1021	Db	CTTGGCTTTCGGGAACCTGGGGCCCTGCTTTCGCCGGCTGCACACAGTGTGCTGCCGCAT	1080
1084	QY	GCCCGCACCCCTGGCGCCGCTCTTTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGAC	1143
1081	Db		
1081	Db	GCCCGCACCCCTGGCGCCGCTCTTTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGAC	1140
1144	QY	CTTCACGCTGTTTTACAGGATTTCTGGCGAGGGCTGTACACAGGGCTGCCAGAGC	1203
1141	Db		
1141	Db	CTTCACGCTGTTTTACACGATTTCTGTGGCGAGGGCTGTACACAGGGCTGCCAGAGC	1200
1204	QY	TGAGCCGGGACCGAGGCCGGAGACACTATGATGA-----	1239
1201	Db		
1201	Db	TGAGCCGGGACCGAGGCCGGAGACACTATGATGAAGGTAAGGCTTTGGCAGCCAGAG	1260
1240	QY	-----	1239
1261	Db	AGGCTGSGTGGAGCCGCCCAACAGAGACACACTCGGGGCTGTGCTGGGCTGGTGCC	1320
1240	QY	-----	1239
1321	Db	TCTCCATCCTGGCCCGACCTCTCTGTTCAGAAAGTGGGATGGACCCCATCTGCATACA	1380
1240	QY	-----	1239
1381	Db	CGGCTTCTCATGGGTGTGGAACATCTCTGCTTCGGGTTTCAGGAAGGCTCTGGCTGCTC	1440
1240	QY	-----	1239
1441	Db	TAGAGTCTCATCAGAGTCGTTGCCCAAGTTTGACAGAAGAAAGCGGAGCTTAATCAA	1500
1240	QY	-----	1239
1501	Db	AGTCTAGAGGGAGTGGAGGAGTTAAGGCTGGATTTTCAGATCTGCCTGGTTCCAGCGGAG	1560
1240	QY	-----	1239
1561	Db	TGTGCCCTCTGCTCCCCCAACGACTTTCCAAAATAATCTCACCAGGGCCTTCCAGCTCAGG	1620
1240	QY	-----	1239
1621	Db	CGTCTTAGAAGCGTCTTGAAGCCTATGGCCAGTGTCTTTGTGTTCCTCTCACC CGCT	1680
1240	QY	-----	1239
1681	Db	GTCCCTCACAGCTGAGACTCCCAGGAAACCTTCAGACTACCTTCCCTCTGCTTCAGCAAGG	1740
1240	QY	-----	1239
1741	Db	GGCGTTGCCCATCTCTCAGGGTCAGTGGGAACCTAGACTCCCATTTGCTAGAGGTA	1800
1240	QY	-----	1239
1801	Db	GAAAGGGAAGGTCGTGGGAGCAGGGCTGGTCCACACAGAGGCTCTCGTGACGAGGTAC	1860
1240	QY	-----	1239
1861	Db	CTGTGGTTCCGGCTTCTCATCTCCCTGAGACTGCTCCGACCCCTCCCTCCAGGCTCTGT	1920
1240	QY	-----AGCGTTCCGATGGCAGGCTGGGGCTGTTCCTGTCAGT	1277
1921	Db		
1921	Db	CTGATGGCCCTCTCCCTCTGCAGGCGTTCCGATGGCAGGCTGGGGCTGTTCCTGCACT	1980
1278	QY	GGCCCATCTCCCTGGTCTCTCTGTGTCATGACCGGCTGGTGACGGATTCGGCACTC	1337
1981	Db		
1981	Db	GGCCCATCTCCCTGGTCTCTCTGTGTCATGACCGGCTGGTGACGGATTCGGCACTC	2040

QY	1338	GAGCAGTCTATTGTTGGCCAGTGTGCAGCTTTCCCTGTGGCTGCCGGTGCACATGCCTGT	1397
Db	2041	GAGCAGTCTATTGTTGGCCAGTGTGCAGCTTTCCCTGTGGCTGCCGGTGCACATGCCTGT	2100
QY	1398	CCACAGTGTGGCCGTGTGTGACAGCTTACAGCCGCCCTCACCGGGTTCACCTTCTCAGCCC	1457
Db	2101	CCACAGTGTGGCCGTGTGTGACAGCTTACAGCCGCCCTCACCGGGTTCACCTTCTCAGCCC	2160
QY	1458	TGCAGATCCTGCCCTACACACTTGCCTCCCTCTACCACCGGGAGAGCAGGTGTTCCTGCG	1517
Db	2161	TGCAGATCCTGCCCTACACACTTGCCTCCCTCTACCACCGGGAGAGCAGGTGTTCCTGCG	2220
QY	1518	CCAAATACCGAGGGGACACTTGGAGGTGCTAGCAGTGAAGACACCTGATGACAGTTCC	1577
Db	2221	CCAAATACCGAGGGGACACTTGGAGGTGCTAGCAGTGAAGACACCTGATGACAGTTCC	2280
QY	1578	TGCCAGGCCCTAAGCCTTGGAGCTCCCTTCCCTAATGGACAGTGGGTGCTGGAGGCAGTG	1637
Db	2281	TGCCAGGCCCTAAGCCTTGGAGCTCCCTTCCCTAATGGACAGTGGGTGCTGGAGGCAGTG	2340
QY	1638	GCCTGCTCCACACTTCCACCCCGCGCTGTCCGGGGCTGTGCTGTGATGTCTCCGTACGCG	1697
Db	2341	GCCTGCTCCACACTTCCACCCCGCGCTGTCCGGGGCTGTGCTGTGATGTCTCCGTACGCG	2400
QY	1698	TGGTGGTGGGTGAGCCACCCAGGCCAGGTGTTCGGGGCGGGGCATCTGCCCTGGACC	1757
Db	2401	TGGTGGTGGGTGAGCCACCCAGGCCAGGTGTTCGGGGCGGGGCATCTGCCCTGGACC	2460
QY	1758	TGCGCATCTCGGATAGTGCTTCTGCTGTCCAGGTGCCCCATCCCTGTTTATGGGCT	1817
Db	2461	TGCGCATCTCGGATAGTGCTTCTGCTGTCCAGGTGCCCCATCCCTGTTTATGGGCT	2520
QY	1818	CCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGTGTCTGCCGAGGCCCTGGGTC	1877
Db	2521	CCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGTGTCTGCCGAGGCCCTGGGTC	2580
QY	1878	TGGTCGCCATTTACTTTGTCTACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACT	1937
Db	2581	TGGTCGCCATTTACTTTGTCTACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACT	2640
QY	1938	CAGCGTAGAAAACTTCCAGCACATTTGGGTGGAGGGCTGCCCTACTGGGTCCCAGCTCC	1997
Db	2641	CAGCGTAGAAAACTTCCAGCACATTTGGGTGGAGGGCTGCCCTACTGGGTCCCAGCTCC	2700
QY	1998	CCGCTCTCTGTAGCCCATGGGGCTCCCGGGCTGCGCCAGTGTCTGTGTGTCGCAAAG	2057
Db	2701	CCGCTCTCTGTAGCCCATGGGGCTCCCGGGCTGCGCCAGTGTCTGTGTGTCGCAAAG	2760
QY	2058	TAATGTGGCTCTGTGTGCCACCTGTCTGTCTGAGGTGCGTAGCTGCACAGCTGGGGGC	2117
Db	2761	TAATGTGGCTCTGTGTGCCACCTGTCTGTCTGAGGTGCGTAGCTGCACAGCTGGGGGC	2820
QY	2118	TGGGGCGTCCCTCTCTCTCCCAAGTCTTAGGGCTGCCTGACTGGAGGCTTTCGAAG	2177
Db	2821	TGGGGCGTCCCTCTCTCTCCCAAGTCTTAGGGCTGCCTGACTGGAGGCTTTCGAAG	2880
QY	2178	GGGGTTTCACTCTGGACTTATACAGGAGGCCAGAGGGCTCCATPGCACTGGAATGCGGG	2237
Db	2881	GGGGTTTCACTCTGGACTTATACAGGAGGCCAGAGGGCTCCATPGCACTGGAATGCGGG	2940
QY	2238	GACTCTGCAGGTGGATTACCCAGGCTCAGGGTTTAACAGTACGCTCCCTAGTTGAGACACA	2297
Db	2941	GACTCTGCAGGTGGATTACCCAGGCTCAGGGTTTAACAGTACGCTCCCTAGTTGAGACACA	3000
QY	2298	CCTAGAGAAGGGTTTTTGGGAGCTGAATAACTCAGTCACTCGTGTGTTCCCATCTCTAAGC	2357
Db	3001	CCTAGAGAAGGGTTTTTGGGAGCTGAATAACTCAGTCACTCGTGTGTTCCCATCTCTAAGC	3060
QY	2358	CCCTTAACCTGCAGCTTTCGTTTAATGTAGTCTTGCATGGCAGTTTCTTAGATGAACAC	2417
Db	3061	CCCTTAACCTGCAGCTTTCGTTTAATGTAGTCTTGCATGGGAGTTTCTTAGATGAACAC	3120
QY	2418	TCCTCATGGGATTTGAACATATG - ACTTAATTTGTAGGGGAAGAGTCTCTGAGGGGCAAC	2475

QY	1241	-----	1240
Db	661	CCAGCTCAGCGCTCCTAGAACGCTCTTGAAGCCTATGGCCAGCTGTCTTTGTGTCCCTC	720
QY	1241	-----	1240
Db	721	TCACCGCCTGTCTCACAGCTGAGACTCCAGGAAACCTTCAGACTACCTTCCCTCTGCC	780
QY	1241	-----	1240
Db	781	TTACGCAAGGGGCTTGCACACATCTCTGAGGGGCTTCGATGGCAGCCTGGGGCTGT	840
QY	1269	TTCTGACAGCGCCATCTCCCTGGCTCTCTCTGTCATGACCGGCTGTGTCAGCCAT	1328
Db	841	TCTGTCAGTGGCCATCTCCCTGGCTCTCTCTGTCATGACCGGCTGTGTCAGCCAT	900
QY	1329	TGGGCACTGACAGCTCTATTTGGCCAGTGTGGCAGCTTTCCTGTGCTGCCGTGCCA	1388
Db	901	TGGGCACTGACAGCTCTATTTGGCCAGTGTGGCAGCTTTCCTGTGCTGCCGTGCCA	960
QY	1389	CATGCTGTCCACAGTGTGGCGCTGTGACAGCTTACGCCCTCCCTTACCCCGGGAGAGCAGG	1448
Db	961	CATGCTGTCCACAGTGTGGCGCTGTGACAGCTTACGCCCTCCCTTACCCCGGGAGAGCAGG	1020
QY	1449	TCTAGCCTGTGAGATCTCTGCTTACACACTGGCCCTCCCTTACCCCGGGAGAGCAGG	1508
Db	1021	TCTAGCCTGTGAGATCTCTGCTTACACACTGGCCCTCCCTTACCCCGGGAGAGCAGG	1080
QY	1509	TGTTCTGCTCCCAATACCGAGGGGACACTGGAGTGTCTAGCAGTGGAGCAGCTGATGA	1568
Db	1081	TGTTCTGCTCCCAATACCGAGGGGACACTGGAGTGTCTAGCAGTGGAGCAGCTGATGA	1140
QY	1569	CCAGCTTCTGTCAGCGCCCTAAGCCTGGAGCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT	1628
Db	1141	CCAGCTTCTGTCAGCGCCCTAAGCCTGGAGCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT	1200
QY	1629	GAGGAGTGGCTGTCTCCACCTCCACCCGGCTGTGCGGGGCTCTGCTGTGATGTCT	1688
Db	1201	GAGGAGTGGCTGTCTCCACCTCCACCCGGCTGTGCGGGGCTCTGCTGTGATGTCT	1260
QY	1689	CGGTACGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1748
Db	1261	CGGTACGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1320
QY	1749	GCCTGGACCTCGCCATCTCTGATAGTGTCTTCTGCTGTCCAGGTGGCCCTCCCTGT	1808
Db	1321	GCCTGGACCTCGCCATCTCTGATAGTGTCTTCTGCTGTCCAGGTGGCCCTCCCTGT	1380
QY	1809	TTATGGGCTCCATTTCTCAGCTCAGCAGTCTGTCTGCTGTCCAGGTGGCCCTCCCTGT	1868
Db	1381	TTATGGGCTCCATTTCTCAGCTCAGCAGTCTGTCTGCTGTCCAGGTGGCCCTCCCTGT	1440
QY	1869	GCCTGGGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1928
Db	1441	GCCTGGGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1500
QY	1929	CCAAATCTCAGCTAGGAACTTCCAGCATTGGGGTGGAGGCTGTGCTTCTACTGGGT	1988
Db	1501	CCAAATCTCAGCTAGGAACTTCCAGCATTGGGGTGGAGGCTGTGCTTCTACTGGGT	1560
QY	1989	CCCAGCTCCCGCTCTCTTGTAGCCCATGGGGCTGGCGGCTGGCGGCTGGCGGCTGGCGG	2048
Db	1561	CCCAGCTCCCGCTCTCTTGTAGCCCATGGGGCTGGCGGCTGGCGGCTGGCGGCTGGCGG	1620
QY	2049	GTGCCAAAGTATGTGGTCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2108
Db	1621	GTGCCAAAGTATGTGGTCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1680
QY	2109	GCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG	2168
Db	1681	GCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG	1740
QY	2169	CCTTCCAAAGGGGTTTTCAGTCTGACTTATACAGGAGGCCAGAGGCTCCATGCACTG	2228
Db	1741	CCTTCCAAAGGGGTTTTCAGTCTGACTTATACAGGAGGCCAGAGGCTCCATGCACTG	1800
QY	2229	GAATCGGGGACTCTGACAGTGGATTACCCAGGCTCAGGGTTAAACAGTACGCTCCCTAGT	2288
Db	1801	GAATCGGGGACTCTGACAGTGGATTACCCAGGCTCAGGGTTAAACAGTACGCTCCCTAGT	1860
QY	2289	TGAGACACACCTAGAGAGGGTTTGGAGCTCAATAAAGTACCTCAGCTGGTTCCTCCCA	2348
Db	1861	TGAGACACACCTAGAGAGGGTTTGGAGCTCAATAAAGTACCTCAGCTGGTTCCTCCCA	1920
QY	2349	TCTCTAAGCCCTTAACTTCAGCTTCTGTTTAAATGTAGCTCTTTCATGGAGTTTCTAGG	2408
Db	1921	TCTCTAAGCCCTTAACTTCAGCTTCTGTTTAAATGTAGCTCTTTCATGGAGTTTCTAGG	1980
QY	2409	ATGAACACACTCTCCATGGGATTTGAACATATG--ACTTATTTAGGGAGAGTCTCTG	2466
Db	1981	ATGAACACACTCTCCATGGGATTTGAACATATG--ACTTATTTAGGGAGAGTCTCTG	2040
QY	2467	AGGGCAACACAAAGAACAGGTCCTTCAGCCACAGACACTGTCTTTTGTGTATCCCA	2526
Db	2041	AGGGCAACACAAAGAACAGGTCCTTCAGCCACAGACACTGTCTTTTGTGTATCCCA	2100
QY	2527	CCCCCTCTTACCTTTTATCAGGATGTGGCTGTGCTCTCTGTTGGTCCCTTCTGTTGCCAT	2586
Db	2101	CCCCCTCTTACCTTTTATCAGGATGTGGCTGTGCTCTCTGTTGGTCCCTTCTGTTGCCAT	2160
QY	2587	CACAGGCAATTAATATTTAACTTATTTAACTTATTTAACTTATTTAACTTATTTAACTT	2646
Db	2161	CACAGGCAATTAATATTTAACTTATTTAACTTATTTAACTTATTTAACTTATTTAACTT	2220
QY	2647	CTTTCTGT	2706
Db	2221	CTTTCTGT	2280
QY	2707	CTGAGT	2766
Db	2281	CTGAGT	2340
QY	2767	CCCCAAATGCCTAACCCAGGACCTTGGAAATTTCTACTCATCCCAATGATATTCCTT	2826
Db	2341	CCCCAAATGCCTAACCCAGGACCTTGGAAATTTCTACTCATCCCAATGATATTCCTT	2400
QY	2827	TGCTGT	2886
Db	2401	TGCTGT	2460
QY	2887	GGCTTCCCTAACCCAGGACCTTCTCTTGTGGCCAGCCTGTGTTCCCCCACTTCCACCTCC	2946
Db	2461	GGCTTCCCTAACCCAGGACCTTCTCTTGTGGCCAGCCTGTGTTCCCCCACTTCCACCTCC	2520
QY	2947	TCTACTCTCTTAGGACTGGGCTGTGAGGCACTGCCAAAATTTCCCTTCCCTTCCCTTCC	3006
Db	2521	TCTACTCTCTTAGGACTGGGCTGTGAGGCACTGCCAAAATTTCCCTTCCCTTCCCTTCC	2580
QY	3007	TTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC	3066
Db	2581	TTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC	2640
QY	3067	CCAGAGCAAAAGTGGGTTTCCCAAGCTTTTGTCCATCTCAGCCCCCAGAGTATATCT	3126
Db	2641	CCAGAGCAAAAGTGGGTTTCCCAAGCTTTTGTCCATCTCAGCCCCCAGAGTATATCT	2700
QY	3127	GTCTTGGGGAATCTCACAGAACTCAGGAGCACCCTTGCCTGAGTAAAGGAGTCT	3186
Db	2701	GTCTTGGGGAATCTCACAGAACTCAGGAGCACCCTTGCCTGAGTAAAGGAGTCT	2760
QY	3187	TTATCTCTCAGGGGGGTTTAAAGTGGCTTTTGAATATGCTCTTATTTTATTTAGCGG	3246
Db	2761	TTATCTCTCAGGGGGGTTTAAAGTGGCTTTTGAATATGCTCTTATTTTATTTAGCGG	2820
QY	3247	GGTGAATATTTTATCTGTAAGTGACCAATCAGAGTATATGTTTATGTTGACAAAATTA	3306

Db 2821 GGTGAATATTTATACCTAAGTGAGCAATCAGAGTATAATGTTATGGTGACAAAATTA 2880
Qy 3307 AAGGCTTTCTATATGTTTAAAAA:3330
Db 2881 AAGGCTTTCTATATGTTTAAAAA 2904

RESULT 10
US-09-605-785-702
; Sequence 702, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605/785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 702
; LENGTH: 4894
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-605-785-702

Query Match 62.8%; Score 2142.8; DB 4; Length 4894;
Best Local Similarity 80.1%; Pred. No. 0;
Matches 2872; Conservative 1; Mismatches 3; Indels 708; Gaps 2;

Qy 455 GGCATTGGTCCAGTCTGGGCGCTGTGTGTGTCGCCGCTCCTAGGCTCAGCCAGTGACCAC 514
Db 1311 GGCATTGGTCCAGTCTGGGCGCTGTGTGTGTCGCCGCTCCTAGGCTCAGCCAGTGACCAC 1370

Qy 515 TGGCGTGGAGCGCTATGGCGCGCCGCCGCTTTCATCTGGGCACTGTCTTGGGCACTCCTG 574
Db 1371 TGGCGTGGAGCGCTATGGCGCGCCGCCGCTTTCATCTGGGCACTGTCTTGGGCACTCCTG 1430

Qy 575 CTGAGCTCTTTCTCATCCCAAGGCGCGCTGGCTAGCAGGCGTGTGTCGCCGATCCC 634
Db 1431 CTGAGCTCTTTCTCATCCCAAGGCGCGCTGGCTAGCAGGCGTGTGTCGCCGATCCC 1490

Qy 635 AGGCGCCCTGGAGCTGGGCACTGTCTATCCTCGGCGTGGGCTGTCTGGGCACTGTCTG 694
Db 1491 AGGCGCCCTGGAGCTGGGCACTGTCTATCCTCGGCGTGGGCTGTCTGGGCACTGTCTG 1550

Qy 695 GTGTGCTTCACTCCACTGGAGCGCTGTCTCTGAGCTCTTTCGGGCACTGTCTGAGCTGT 754
Db 1551 GTGTGCTTCACTCCACTGGAGCGCTGTCTCTGAGCTCTTTCGGGCACTGTCTGAGCTGT 1610

Qy 755 CGCAGCGCTTACTGTCTGTATGCTTCATGATCAGTCTTGGGCGCTGCTGGGCACTGTCTG 814
Db 1611 CGCAGCGCTTACTGTCTGTATGCTTCATGATCAGTCTTGGGCGCTGCTGGGCACTGTCTG 1670

Qy 815 CTGCTGCCATTGACTGGGACACAGTGCGCTTGGCGCTTACCTGGGCACTCCAGGAGGAG 874

Db 1671 CTGCTGCCATTGACTGGGACACCACTGCTGCCCTGCTGCCCTTACTCTGGGACCCAGGAGGAG 1730
Qy 875 TGCTCTTTTGGCTGTCTACCCCTCATCTTCTCCTCAGCTGGGTAGCAGCCACACTGCTGGTG 934
Db 1731 TGCTCTTTTGGCTGTCTACCCCTCATCTTCTCCTCAGCTGGGTAGCAGCCACACTGCTGGTG 1790

Qy 935 GCTGAGGAGGACGGCTGGGCGCCACCGAGGCGACAGAGGGCTGTGGGCGCCCTCTCTTG 994
Db 1791 GCTGAGGAGGACGGCTGGGCGCCACCGAGGCGACAGAGGGCTGTGGGCGCCCTCTCTTG 1850

Qy 995 TCGCCCACTGCTGTCCATCGCGGCGCTTTCGGGAACTTTCGGGCGCCCTGCTT 1054
Db 1851 TCGCCCACTGCTGTCCATCGCGGCGCTTTCGGGAACTTTCGGGCGCCCTGCTT 1910

Qy 1055 CCGCGGCTGCACGAGCTGTCTGCCGATGCCCGCACCCCTGCGCGGCTTCTTCGTTGGCT 1114
Db 1911 CCGCGGCTGCACGAGCTGTCTGCCGATGCCCGCACCCCTGCGCGGCTTCTTCGTTGGCT 1970

Qy 1115 GAGCTGTGACGCTGGATGGGCACTCATGACCTTCACGCTGTTTACAGGATTTCTGTTGGC 1174
Db 1971 GAGCTGTGACGCTGGATGGGCACTCATGACCTTCACGCTGTTTACAGGATTTCTGTTGGC 2030

Qy 1175 GAGGCGCTGTACCAGGCGCTGCCAGAGCTGAGCGCGGCACCCGAGGCGCCGAGACACTAT 1234
Db 2031 GAGGCGCTGTACCAGGCGCTGCCAGAGCTGAGCGCGGCACCCGAGGCGCCGAGACACTAT 2090

Qy 1235 GATGA----- 1239
Db 2091 GATGAAGGTAAGGCTTTGGCAGCAGCAGAGGCTGTGTGGGAGCGCCGCCAGAGACG 2150

Qy 1240----- 1239
Db 2151 AACTCGGGGCTGTGTGTGGCTGGTGCCTCTCCATCCTGCGCCCGACTTCTCTGTGAGG 2210

Qy 1240----- 1239
Db 2211 AAGTGGGATGAGACCCCATCTGCATACACGGCTTCTCATGGGTGTGGAACATCTCTGCT 2270

Qy 1240----- 1239
Db 2271 TGGGTTTTCAGGAAGGCGCTGTGGCTGTCTAGGAGTCTGATCAGAGTGTGTCCTCCAGTT 2330

Qy 1240----- 1239
Db 2331 TGACAGAGGAAAGCGGAGGCTTATCAAAGTCTAGAGGAGTGAGGAGTTAAGGCTGG 2390

Qy 1240----- 1239
Db 2391 ATTTTCAGATCTGCTGTGTTCAGCGCGAGTGTGCCCTCTGCTCCCGCAAGACTTTCCAA 2450

Qy 1240----- 1239
Db 2451 ATAACTCACCAGCGCTTCCAGCTCAGGCGTCTAGAGGCTTGAAGGCTATGGCCA 2510

Qy 1240----- 1239
Db 2511 GCTGCTTTGTGTTCCCTCTCACCGCGCTGTCTCACAGCTGAGACTCCCGAGAAACCTT 2570

Qy 1240----- 1239
Db 2571 CAGACTACCTTCTCTGTGCTTTCAGCAAGGGGCGTGTGCCACATTTCTCTGAGGCTCAGTGG 2630

Qy 1240----- 1239
Db 2631 AAGAACCTAGACTCCCATTTGCTAGAGGTAGAAAGGGAAGGCTGTGGGAGCAGGCGTG 2690

Qy 1240----- 1239
Db 2691 GTCCACAGCAGGCTCTGTCAGCAGGTAACCTGTGTGTTCCGCTTCTCATCTCCCTGAGAC 2750

Qy 1240----- 1248
Db 2751 TGCTCCGACCTTCCCTCCAGGCTCTGTCTGATGGGCGCTTCCCTCTGAGGCGGCTCG 2810

Result No.	Query			DB	ID	Description
	Score	Match	Length			
C 1	398.5	6.4	660	1	QBEB3	BHLF1 protein - hu
C 2	373	6.0	1453	2	S21626	collagen alpha 1(I)
C 3	364	5.8	1414	1	S23809	collagen alpha 2(I)
C 4	358.5	5.8	1464	2	S29856	collagen alpha 1(I)
C 5	351.5	5.6	1466	1	CGH07L	collagen alpha 1(I)
C 6	348	5.6	1690	1	CGHUIB	collagen alpha 4(I)
C 7	347.5	5.4	515	2	T14340	collagen-proton tra
C 8	344.5	5.5	1049	1	CGB07S	collagen alpha 1(I)
C 9	343	5.5	1464	1	CGHUIS	collagen alpha 1(I)
C 10	342	5.5	1042	1	SCGHUS	collagen alpha 1(I)
C 11	337.5	5.3	523	2	T12198	collagen alpha 1(I)
C 12	335.5	5.4	886	2	S150694	sucrose transport
C 13	335.5	5.4	1549	2	T48103	collagen alpha 1(I)
C 14	335.5	5.4	1843	2	S18803	type VII collagen
C 15	333.5	5.4	15803	2	S18803	collagen alpha 1(V)

```
QY 2183 AACCCCTTGGAGGCTTCCAGTCCAGCCCTAGAGACTGGGAGAGAGAGGAC 2124
Db 31 GlnThrArgArgLeuProPro-----GlyTrpGlyGlnArgThr 43
QY 2123 GCCCGAGCCCGAGCTGTGACGTACGACCTCAGCAGCAGAGGCTGGCAGCAGAGCC 2064
Db 44 AlaProThrGlnValGlyLeuAlaAspAlaSerProAspGluLeuGlnAspGln--- 62
QY 2063 ACATTACTTTGGCAGCAACAGAAACTGGCGCCAGCCCGCAGCCCTCGGCTATACAG 2004
Db 63 -----AlaSerGlyAlaArgProGlyGlyGlyAsn 72
QY 2003 GAGCGGAGCTGGGACCCAGTGGAGGAGCCCTCCACCCCAATGTCTGGAAGTTTCT 1944
Db 73 ArgValGlyAlaGlyArg-----GlyArgProGlyThrProAla----- 85
QY 1943 ACCTGTAGTATTGGCCAAAGTCGCTTGTCAATACTACTGTGTAGCAAGTAAATGG 1884
Db 86 -----ProSerArg-----GlnSerArg--- 91
QY 1883 GCACAGACCCA-----GGCCTCGGCAG 1860
Db 92 ArgThrGlyProAlaGluGlnAlaAspHisAlaHisSerAsnProThrGlyGlyCysSer 111
QY 1859 ACACCATATAGCAGTACAGACTGGCTGAGCTGGACAAATGGAGCCCAATAACAGGATG 1800
Db 112 AspProGlnArgSerProArgThrArgGlnAlaGlyTy-----AlaLeu 126
QY 1799 GGGCACCCTGGGACAGCAGGAGGACATATCCAGATGGCAGTGCCA----- 1752
Db 127 Gly---GluGlySerAlaGlyLeu-----GlySerArgGlyProArgProHisPro 142
QY 1751 -----GGCAGATGCCCGCCCGGCAACACCTGCTCGGTGG 1713
Db 143 AlaPheGlnValGlnTrpSerAlaArgAsnProGlyCysProArgThrTrpArgArg 162
QY 1712 GCT-----CACCCACCACACGCTACGAGACATCAGCGCAGGAGGCC 1668
Db 163 SerGlyAlaGlnArgGlyHisProPro----- 172
QY 1667 CCAGAGCCGGGTGGAGGTGGAGCAGGCCACTGCTCCAGCACCACGCTGTCATTAG 1608
Db 173 -----GlyAlaGly----- 175
QY 1607 GGAAGGAGCTCCAGGCTTAGGGCTTGGCAGGAGCTGGTCATCAGGCTGCTCCTACTGC 1548
Db 175 ----- 175
QY 1547 TAGCAGCTCCAGTGTCCCTCGTATTTGGGAGGAGAACACCTGCTTCCCGGTGAGTA 1498
Db 176 -----GlnArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaPro 192
QY 1487 GGGAGGCCAGTGTAGGCGCAGGATCTCAGGCGCTGAGAAGGTGAACCGGTGAGGCGG 1428
Db 193 GlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProSer----- 208
QY 1427 CTGAAGCTGTACACCGCCACACTGTGGGACAGC-----ATGTGG 1386
Db 209 ---GlyAlaThrProHisProGluArgGlySerGlyProAlaAspProAlaAlaAla 227
QY 1385 CACCGCAGCCAGGGAAGCTGGCAGACTGGCCCAATAGACTCTCGAGTGGCGAATC 1326
Db 228 ArgLeuProGluArgGlnGluProArgLeuProGlnAspLeuAlaAlaGln--- 246
QY 1325 GCTGCACACCGGCTCCATGACAGAGAGAA-----CCAGGAGATGGCCACTCCAGGA 1269
Db 247 ArgCysProAlaGlyProProProProThrArgSerGlyAlaAlaAlaGlnArgHisArg 266
QY 1268 ACAGCCCGAGGCTGCCATCCGAACGCTTCATATAGTGTCTCCGGGCTCGGTGCCG 1209
Db 267 ArgProProGlyCysPro-----ArgSerAlaArgAsnProGlyCysPro 281
QY 1208 GCTCAGCTCTGGCAGCAGCCCTGTAGAGCCCTCCGCCACGAAATCCGTGTAAACACGG 1149
Db 282 ---ArgThrTrpArgArgSerGlyAlaGlnArgGlyHisProPro----- 296
QY 1148 TGAAGTCTATGAGTCCATCCAGCTCAGCCAGAGAGCGCGCAGGAGTGC 1089
Db 297 -----ProGlyAlaGlyGln 301
QY 1088 GGGCATCGGCGCAGCAGCTGTGTGAGCCCGGGAAGCAGGGCGCCAGTTCCGGAAG 1029
Db 302 ArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrProAla 321
QY 1028 CCAAGCGCCCGCGCATGACAGC-----AGTGGGCGCACA-----AGGAGG 987
Db 322 AlaProGlyProGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArg 341
QY 986 GGG-----CCGACAGCCCTTCTGCTG-----GCTCGGTGGGCGCCA 951
Db 342 GlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGluPro 361
QY 950 GCCTGCTCTCTCAG-----CCACGACGAGTGTGG 921
Db 362 ArgLeuProGlnAspLeuAlaAlaGlnArgCysProAlaGlyProProProThr--- 380
QY 920 CTCTACGCGAGTGAAGATGAGGTGAGCAGGCGCCAAAGA-----GGCCTCTCTCT 867
Db 381 -----ArgSerGlyAlaAlaGlnArgThrHisArgArgProPro 394
QY 866 GGCTGCCCA-----GCTAGGGCGCCAGGCGCTGTGCTCCAGTCAATGGCAGGCGAGA 813
Db 395 GlyCysProArgSerAlaArgAsnProGly-----CysProArgThrTrpArg----- 410
QY 812 GGTAGCCAGGCGCCCAAGACTGATCATGAAGCATAGACAGAGTGGCTGCGGAC 753
Db 411 -----ArgArgSerGlyAlaGlnArg 417
QY 752 AGTGTGCTGGTCCCGAGAGGCTCAGAGA-----GCAGGCGCTCCAGTG 708
Db 418 GlyHisProProProGlyAlaGlyGlnArgProSerGlyProThrGlyArgProAla 437
QY 707 GAGTGAAGCAGCTGGCCACAGAAAGTCCAGCAGCCCGCCAGCGCAGGATGAGCAGTGCCA 648
Db 438 AlaProGlyAlaProGly-----ThrProAlaAlaProGlyProGlyGlyAlaAla 455
QY 647 GCTCCA---GGGCGCTG-----GGATCCGGCGCAGCAGCAGCCCTGCTAGC 607
Db 456 ValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProProAla 475
QY 606 CAGCGCGCTCTGGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGAC----- 559
Db 476 AlaAlaArgLeuProProGluArgGlnGluProArgLeuProGlnAspLeuAlaAla 495
QY 558 -----AGTCCCGCAGATGAAGGC 541
Db 496 GlnArgCysProAlaGlyProProProThrArgSerGlyAlaAlaGlnArgThrHis 515
QY 540 CGGCGCGCGCATAGCTCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 481
Db 516 ArgArgProPro-----GlyCysProArgSer---AlaArgAs 527
QY 480 ACCAGGCGCCAGCAGCTGAGCAATGCCAGCAGCAGCATGGTTCATGCTTCTCTCTCTCT 426
Db 527 nProGly-----CysProArgThrTrpArgArgSerGlyAlaGlnArg 542
QY 425 -----CCCCCACTTCCAGCAGCAGAGGCGCCACATAGGTGATGCTTCTCTCTCTCT 373
Db 542 gGlyHisProProProGlyAlaGlyGln-----ArgProSerG 555
QY 372 ACCTCCAGGCGCAAGGTTAGCAGGTTCACAGCAGAGAGCTGGCTTTTCCGCTGCGCGCAG 313
Db 555 yProThrGlyGlyArg-----ProAlaAlaProGly-----Al 566
QY 312 AGGCGGTTCACCCACAGCCTCTGTGACCATAGTGGCGCCAGGCGGTAGGCTCAGGGGCGC 253
```


Db 209 gGlyProProGlyProPro 215
1069 AGAGCCACATTTACTTTGGCAGCAACAGAAACTGGCGCCAGCCGAGCCAGCCCATGGGGC 2010
216 -----GlyLysAsnGlyAspAspGlyGluAlaGlyLysProGlyArgPr 230
2009 TAACAGAGCGGGAGCTGGGACCCAGTGGAGCCAGGAGCCCTCCACCCCAATGTCTGGAAG 1950
230 oGlyGluArgGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGly-- 249
1949 TTTTCTAGCTGAGTATTGGCCAAAGTCGCTCTGTCAATATACCTGTGTACCAAGT 1890
250 -----LeuProGlyMetLysGlyHisArgLysPheSerGlyLeuAspGlyAl 265
1889 AAATGGCAGCAGACCCAGGCTCGGCAGCAGACCATATAGGCAGTGA-----CAGACTGG 1834
265 aLysGlyAspAlaGlyProAlaGlyProLysGlyGluProGlySerProGlyGluAsnG 285
1833 CTGAGCTGGCAANTGGAGCCCAATACAGGGATG-----GGGCC 1795
285 yAlaProGlyGlnMetGlyPro--ArgGlyLeuProGlyGluArgGlyArgProGlyPr 304
1794 ACCTGGGACAGGAGGCACTATCCAGGATGCGAGGTCCAGGCAGATGCCCGGCC 1735
304 oProGlyThrAlaGlyAlaArgGly-AsnAspGlyAlaValGlyAlaAlaGlyPro-Pro 323
1734 GGAACCCCTGGCTCGGTGGCTCACCACCCAGCAGTACGAGACATCAGGCA 1675
324 GlyProThr-----GlyProThrGlyProProGlyPheProGlyAlaValGlyAlaLysGly 342
1674 GAGGCC-----CGCAGAGCGCGGTGGAGGT-----GGGAGCAGG-----CCA 1636
343 GluAlaGlyProGlnGlyAlaArgGlySerGluGlyProGlnGlyValArgGlyGluPro 362
1635 CTGCTCCAGCAGCCAGCTGCTCCATTAGG-----RAGGAGCTCCAGCTTAGG----- 1585
363 GlyProProGlyProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGln 382
1584 CTGGCAGGAAGCTGCTCAGGCTGCTCCTCACTGCTAGCACCTCCAGTCTCCCTCGG 1525
383 ProGlyAlaLys-GlyAlaAsnGlyAlaPro-----G 393
1524 TATTGGGAGGAACACCTGCTTCCCGGTGGTAGAGGGAGCCAGTGTGTAGGGCAGG 1465
393 yIle---AlaGlyAlaProGlyPheProGlyAlaArgGly-----ProSerG 408
1464 ATCTCAGGCTGAGAAAGTGAACCGGTGAGGGCGCTGAAGCTGTCCACACGGCCACA 1405
408 yPro-GlnGlyProSerGlyProProGlyProGlyProLysGlyAsnSerGlyGluProGly---A 427
1404 CTGTGGGACAGGCATGTGGCACCAGCCAGCAGGGAAGCTGCCACACTGGCCCAATAG 1345
427 laProGlyAsnLysGlyAspThrGlyAlaLysGlyGluProGlyAlaThrGly----- 444
1344 ACTGCTCGAGTCCGCAANTCGCTGACCCAGCCCGTCCATGACAGAGAGACAGGAG 1285
445 -----ValGlnGlyPro--ProGlyProAlaGlyGluGluGlyL 457
1284 ATGGCGCACTGCGAGGAACA-----GCCCGAGCTGCCATCCGAACGC--- 1242
457 sArgGlyAlaArgGlyGluProGlyProSerGlyLeuProGlyProProGlyGluArgG 477
1241 -----CTTCATCATAGTGTCTCCGGGCTCGGTGCGCGGC 1207
477 yGlyProGlySerArgGlyPheProGlyAlaAspGlyValAlaGlyProLysGlyProSe 497
1206 TCAGCTCTGGCAGCCCTGTGTACAGCCCTCCGCCACCAATCGGTGTAAACACGCTG 1147
497 rGlyGluArgGlyAlaProProGlyProAlaGlyProLysGlySerPro----- 512
1146 AAGGTCTAGTATGCCATCCAGCTGCACAGCTCAGGCACGAGAGCGCGCGAGGTGCGG 1087

512 ----- 512
1086 GCATCGGCACACAGCTGTGTGACCGGGGAAGCAGGCGCCAGCT----- 1038
513 -----GlyGluAlaGlyArgProGlyGluAlaGlyLeuProGlyAlaLysGlyLe 529
1037 -TCCGAAAGCCAGCGCCGCGCATGACAGCAGTGGGGCAGCAAGAGGGGCCAC 979
529 uThrGlySerProGlySerProGlyProAsp-----GlyLysThrGlyProProGlyPr 547
978 AGCCCTTCTGCTGGCTGGTGGGGCCAGCGCTGCT----- 942
547 oAlaGlyGlnAspGlyArgProGlyProAlaGlyProProGlyAlaArgGlyGlnAlaG 567
941 -----CCTCAGCCACAGCAGTGTGTGCTTACCGAGGTGAGGAAGATGAGG 895
567 yValMetGlyPheProGlyProLysGlyThrAla-----GlyGluProG 582
894 GTGAGCAGCCAAAGAGCAGCTCT-----GGTAGCCAGGAGCCGCCA 793
582 yLysAlaGlyGluArgGlyLeuProGlyProGlyAlaValGlyProAlaGlyGluArgGlyG 622
869 -----CCTGGGTGCCAGGT---AGGGGGCCAGGCGCACTGTGTCCAC 829
602 pGlyGluAlaGlyAlaGlnGlyAlaProGlyProAlaGlyProAlaGlyGluArgGlyG 622
828 TCAATGGCAGGCAAG-----GGTAGCCAGGAGCCGCCA 793
622 uGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyProAlaGlyProG 642
792 AGACTCATCATGAAGCATAGACAGTAGGCTGCGCAGAGTGT---CCGGTCCC--- 738
642 yGluAlaGlyLysProGlyGluGlnGlyValProGlyAspLeuGlyAlaProGlyProSe 662
737 ---GGAAGAGGTGACAGGAGCGCTCCAGTGAGTGAAGCAGCACACCTGGCCAGAG 682
662 rGlyAlaArgGlyGluArg---GlyPheProGlyGluArgGlyValGlnGlyProG 681
681 TCAGCAGCCCGCCAGCCAGGATGAGCAGTGC--- 648
681 yProAlaGlyProArgGlyAsnAsnGlyAlaProGlyAsnAspGlyAlaLysGlyAsp 701
647 ---GCTCCAGGCGCTGGATCCGGGCACAGCAGCCCTCTAGCCAGCGCGCTTGGG 592
701 rGlyAlaProGlyAla-----ProGlySerGlnGlyAlaProGlyLeuGlnGlyMetPr 719
591 ATGAAAGAGCTCAGCAGGATGCCCAAGCAGCAGTCCCATGATGAAGCGCGCGCGG 532
719 oGlyGluArgGlyAlaAlaGlyLeuProGly-----ProLysGlyAspArgGlyAsp 737
531 CCATAGCTCCAGCCAGTGTCTGCTGAGCTGAGCTAGGAGCGGACACAGACAGCCGCC 472
737 aGlyProLysGlyAlaAspGlySerProGlyLysAspGlyAlaArgGlyLeuThrGlyPr 757
471 AGCACTGGACCAATGCGCCAGCAGCAGCATGTCATCACTTCTCTACCCCACTTCCAGC 412
757 oIle-----GlyProProGlyProAl 764
411 AGCAGAGCGGCACATAGGTGATGCTGCGGCCCAACACACACCTCCAGGC----- 363
764 aGlyAlaProGlyAspLysGlyGluAlaGlyProSerGlyProProGlyProThrGlyAl 784
362 -----CAAGGTTAGCAGGTGACACAGAGTGGCTGGCTTCCGG 322
784 aArgGlyAlaProGlyAspArgGlyGluAlaGlyProProGlyProAlaGlyPheAlaG 804
321 T---GCCGACAGCGGCTCACCACAGCCTCT----- 291
804 yProProGlyAlaAspGlyGlnProGlyAlaLysGlyGluProGlyAspThrGlyValL 824
290 -----GGACCATAGTGGGCGAGCGGG---TAGGCTCAGGCGGC 253
824 sGlyAspAlaGlyProProGlyPro-AlaGlyProAlaGlyProProGlyProLleGlyA 844

QY 252 GTTCAGGCACCTCCAGAACTGCTTCTGCTCTGCTCTCCAGAAAGTCGGGCTCTCTC 193
 Db 844 snValGlyAlaProGly-----ProLysGlyProArgGlyAlaAlaGlyProp 860
 QY 192 CTTCGCTGCCGCCCAACTGCTAGCAATCAGCCAGGCGCCCATTTTCGCGAGCCCTTTGGTG 133
 Db 860 ro-GlyAlaThr-----GlyPheProGlyAlaAlaGlyArgValGlyProGly 876
 QY 132 CCGGTCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGGGCGCACCTCAGT----- 79
 Db 877 ProSerGlyAsnAlaGlyProGlyProGlyProGlyProValGlyLysGlyGlyLys 896
 QY 78 -----GGGACACGCTCATCCTCAGTCTGCGCGA----- 46
 Db 897 GlyProArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyProGlyPro 916
 QY 45 -----GGCGCGCGCTGTCACCCGAGCC 22
 Db 917 ProGlyProAlaGlyGluLysGlySerProGlyAla 928

RESULT 3

S23809
 collagen alpha 2(I) chain homolog - sea urchin (Strongylocentrotus purpuratus)
 C:Species: Strongylocentrotus purpuratus (purple urchin)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S23809
 R:Exposito, J.Y.; d'Alessio, M.; Solursh, M.; Ramirez, F.
 J. Biol. Chem. 267, 15559-15562, 1992
 A:Title: Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha-2(I) collagen
 A:Reference number: S23809; PMID:92348411; PMID:1639795
 A:Accession: S23809
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1414 <EXP>
 A:Cross-references: EMBL:M92040; NID:g161435; PIDN:AAA30035.1; PID:g161436
 C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
 C:Keywords: collagen; triple helix
 F:1207-1414:Domain: fibrillar collagen carboxyl-terminal homology <CC>

Alignment Scores:

Procl. No.: 1-91e-14 Length: 1414
 Score: 364.00 Matches: 321
 Percent Similarity: 30.88% Conservative: 57
 Best Local Similarity: 26.23% Mismatches: 381
 Query Match: 5.85% Indels: 467
 DB: 1 Gaps: 78

US-09-759-143-110 (1-3410) x S23809 (1-1414)

QY 3029 GGGAAAGTTGGGTAGGGAAAGTTGGGTAGGGAAATTTTGGCAGTGCCTTCATC 2970
 Db 43 GlyAspLysGlyGlnLysGlyGluProGlyAspAlaAspIle---AsnSerAlaAsnPhe 61
 QY 2969 AGCCAGCTCTAGAGAGAGTAGAGGG-----AGTGGAGTGGG 2931
 Db 62 ProProGlyLeuProGlyProValGlyProGlyProSerGlyProSerGlyProAla 81
 QY 2930 GGGAAACAGGCT-----GGGCCAAGA-----GAAGAGGGGTGG 2898
 Db 82 GlyAsnAsnGlyProGlyProAsnGlyProArgGlyAsnProGlyMetAspGlyLeu 101
 QY 2897 TTAGGAAGCGTTGAGACCTGAAGCCGCCCTCTACCT----- 2858
 Db 102 ThrGlyLeuProGlyIlePro---GlyProProGly-ProProGlyLysSerGlySerLe 120
 QY 2857 -----TCCTTCAACACCTTAACCTTGGTGGTAACACATTTGGAA 2820
 Db 120 uValAlaSerAlaGlnThrSerSerPheAsnLys-----GlyProSerLeuAlaG 137
 QY 2819 TTATCATTTGGGATGAGTAGAATTTTCCAAAGTCTCGGTAGGCATTTTGGGGGCCAGA 2760
 Db 137 yTyr-----GlnTyrProGlnAla-----GlnAlaAlaGlyThrProG 150

QY 2759 CCCAGGAGAGAAGATTCTGGCAATGATCAGCCCCAATGACCAGCTATCTCAGGGACCT 2700
 Db 150 yProArg----- 152
 QY 2699 GATTGTTGGGATCCCCACCCCTACCAAAATATTAGACACCAACACACAGAAAGCTACGAA 2640
 Db 153 -----GlyProProGlyProGlySerArgGlyProGlnGlyLeuThrGlyPr 169
 QY 2639 TGGATTCCCTTCTACTTTGTTAAATAAATAAGTTAAATATTAAATCCCTGCTCTCTGT 2580
 Db 169 oSerGlyProSer-----GlyGluThrGlyPr 178
 QY 2579 GATGCAACAGAGGACCAACAGCCACATCTCTATATAAAGGTAAAGGGGGGTGGATCA 2520
 Db 178 oSerGlyAsnSerGlyProGlyProGlyProGlyProGlyProGlyProGlyProGly 198
 QY 2519 GCAAAAAGACAGTCTGT-----GGGCTGAGGGGACCTGGTTCTTGTGTGGCCCT 2467
 Db 198 yAspAspGlyThrProGlySerGlnGlyGlnArgGlyProAlaGlyThrProGlySerAr 218
 QY 2466 CAGGACTCTTCCCTCAAAATAAGTCAATATGTTCAATCCCTGAGGAGTGTTCATCC 2407
 Db 218 gGlyThrProGlyMetProGlyAlaProGlyMetLysGlyHisGlnGlyLeu-----Pr 236
 QY 2406 TAGAACTCCCATGCAAGAGCTACATTAAACAGAGCTGACAGTTAAGGGCTTAGAGATG 2347
 Db 236 oGlyMetThrGlySerLysGlyGluArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 255
 QY 2346 GGAACACGAGTGTAGTGTATTTACGTCTCCCAAAACCTCTCTAGGTGTGTCTCAAC 2287
 Db 255 LysProGlyProValGlyAlaProGlyProAlaGlyProSerGlyGlnProGlyGluA 275
 QY 2286 TAGGA-----GGCTAGCTGTGTA 2269
 Db 275 rgGlyArgThrGlyProAlaGlySerGlnGlyAspArgGlyAlaAspGly-----AlaT 293
 QY 2268 CCTGAGCCCTGGGTAAATCCA-----CCTGAGAGTCCCG----- 2234
 Db 293 hrGlySerGlnGlyProGlySerThrGlyProAlaGlyAlaProGlyMetProGlyI 313
 QY 2233 -----CATTCAGTGCATGAGCCCTCT-----GGCTCCCTGTATAAGTCC 2191
 Db 313 leSerGlyAlaLysGlyAspAlaGlySerProGlyAlaArgGlySerPro----- 329
 QY 2190 AGACTGAACCCCTTGGAGGCTCCAGTCAGGAGCCCTAGAGCTGGGGAGAGAGG- 2132
 Db 330 -----GlyLeuGlnGlyAlaArg-----GlyGluArgG 339
 QY 2131 -----AGAGGAGCGCCAGCCCGCTGTGACAGTACGC----- 2096
 Db 339 LysGlyGlySerGlnGlyGlnThrGlyProProGlyValProGlyArgAspGlySerA 359
 QY 2095 -----ACCTCAGCAGCAGGCT-----GGCA 2074
 Db 359 snGlyAlaLysGlySerAlaGlyProSerGlyAlaGlnGlyThrProGlyPheProGlyA 379
 QY 2073 GCAGA-----GAGCCACATTACTTTGGCACAACAGAACTGGC 2035
 Db 379 laArgGlyProProGlyProAlaGlySerProGlyProAlaGlySerLysGlyAspGlnG 399
 QY 2034 GGCCAGCCCGGAGCCCGCCATGGGCTTAACAGAGGCGGGAGCTGGGACCC-----AGTGAGG 1978
 Db 399 LysAsnProGlyGlnProGlyAlaGlnGlyGlySerGlyProLeuGlyProArgGlyGluT 419
 QY 1977 CAGGCCCT-----CCACCCCAATGTGTGGAAGTTTCTACGCTGAGTATTGGCCAAAG 1924
 Db 419 hrGlyProAlaGlyPro----- 425
 QY 1923 TCCTCTTGTCAAATACTACCTGTGTAGCAAAAGTAATGGCGACACAGACCCAGGCTGGC 1864
 Db 426 -----GlyAlaGlnGlyGluSerGlyGluArgGly 436

Db 664 lyLys-AspGlySerProGlyGluPro--GlyAlaAsnGlyLeuProGlyAlaAlaGlyG 483
QY 1995 AGCTGGGACCCAGT-----GAGGCAG 1975
Db 483 luArgGlyProSerGlyPheArgGlyProAlaGlyProAsnGlyIleProGlyGluLysG 503
QY 1974 GCCCTCCA-----CCCCAATGTCTGGAAGTTTCTACGCTGAGTATTG 1930
Db 503 lyProProGlyGluArgGlyGlyProGlyProAlaGlyPro-----ArgGlyValAlaG 521
QY 1929 GCCAAGTCGCTTGTCAAAATACCTACCTGTGTAGCAAAAGTAATAGGCGACACCCAGG 1870
Db 521 lyGluProGly-ArgAspGlyThr-Pro-----GlyGlyProGly 533
QY 1869 CCTGGCGGACACCATATAGGCAGTACAGACTGGCTGAGCTGACAAATGGAGCCCAT 1810
Db 533 yIleArgGlyMetProGlySerProGlyGlyProGlyAsnAspGlyLysProGlyProPr 553
QY 1809 AACAGGGATG-----GGCCACCTGGGACAGCAGGAAGGCAC----- 1773
Db 553 oclySerGlnGlyGluSerGlyArgProGlyProProGlyProSerGlyProArgGlyG 573
QY 1772 -----TATCCAGGATGCGAGGTCCAGGACAGATGCCCGCCCGGAC- 1730
Db 573 nProGlyValMetGlyPheProGlyProLysGly-AsnAspGlyAlaProGlyLysAsnG 593
QY 1729 -----CACCTCGCTCGGTGGCTCACCCACCACACACAGTACGAGACATCA 1681
Db 593 lyGluArgGlyGlyProGlyGlyProGlyLeuPro----- 604
QY 1680 CAGGCAGAGGCGCCGAGCGGGTGGAGGTGGAGCAGCGCCACTCCCTCCAGCACCC 1621
Db 605 -----GlyProAlaGlyLys-AsnGlyGluThrGlyProGlnGlyProGlyPro 621
QY 1620 AGCTGTCATTAGG-----AAGGA 1600
Db 622 ThrGlyProAlaGlyAspLysGlyAspSerGlyProProGlyProGlnGlyLeuGlnGly 641
QY 1599 GCTCAGGCTTAGG----- 1585
Db 642 IleProGlyThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLys 661
QY 1584 -----CCTGCAGAGAGCTGGTGCATCAGCTGTCTCTCACTG 1549
Db 662 GlyGluValcylAlaProGlyAlaProGly-GlyLysGlyAspSerGlyAlaProGlyG 681
QY 1548 CTAGCACCTCCAGTGTCCCTCGGTATT-----TGGCAGGAACACCTGC 1504
Db 681 uArgGlyProProGlyThrAlaGlyIleProGlyAlaArgGlyGlyAlaGlyProProG 701
QY 1503 TTCTCCCGGTGTAGAGG-----AGGCCAGTGTGTAGGGCAGGATCTGC 1459
Db 701 yProGluGlyGlyLysGlyProAlaGlyProProGlyProProGlyAlaSerGlySerPr 721
QY 1458 AGGCGTGAAGGTGAACCGGTGAGCGGCGGTGAAGCTGTCCACCAGCCACACTGTGG 1399
Db 721 oGlyLeu-GlnGlyMetProGlyGluArgGly----- 731
QY 1398 GACAGGCATGTGGCAGCCGCGCAGCACAGGGAAGCTGCCACACTGGCCAAATAGACTGCT 1339
Db 732 -----GlyProGlySerProGlyProLysGlyGluLysGlyGluProGlyGlyA 748
QY 1338 CGAGTGCCGAATCGCTGCA-----CCAGCCCGT----- 1311
Db 748 laGlyAlaAspGlyValProGlyLysAspGlyProArgGlyProAlaGlyProIleGlyP 768
QY 1310 --CCATGACAGAGAGAACAGCAGGAGATGGCGCACTCCAGGACAGCCCGCAGCTGC- 1254
Db 768 roProGlyProAlaGlyGlnProGlyAsp---LysGlyGluGlyGlySerProGlyLeuP 787
QY 1253 -----CCATCCGACGCGCTTCATCATAGTGTCTCCGGGC 1220

787 roGlyIleAlaGlyProArgGlyGlyProGlyGluArgGlyGluHis-----GlyP 804
1219 CTGGTGCCTGGCTAGCTCTGGGCAGCGCTGTGTACAGCCCTCGCCACGAATCCGT 1160
804 roProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGluPro----- 820
1159 GTAAACACAGGTGAAGGTATAGTCCCATCCAGCTGCACAGCTCAGCCACCAAGACCG 1100
820 ----- 820
1099 GCCCAGGTGCGGGCATCGGCAGCACAGCTGGTCAGCCGGGAAGCAGGCGGCCCA- 1041
821 -----GlyAlaLysGlyGluArgGlyAlaProG 830
1040 -----GTTCCGAAAGCCAAAGCGGCCCGCATGTGACAGCAGTGGGGCGCAAGA 989
830 lyGluLysGlyGlyGlyProProGlyProAla----- 841
988 GGGGGCGCAGACCCCTCTGCTGGTGGGGCCAGCGTGTCCCTCAGCCACCAG 929
842 --GlyProThr-----GlySerSerGlyProAlaGlyProGlyProGlyProGlnG 957
928 CAGTGTGGCTGCTACGCGAGGTGAGGAAGATGAGGTGAGCAGGCCCAAGAGGACACT- 873
857 lyVal-----LysGlyGluArgGlySerProG 866
872 --CCTCTGGTGGTCCAGGTAGGGGCCAGGCGACTGGTGTCCCATCTAATGCCAGGCAG 815
866 lyProGlyThrAlaGlyPheProGlyGlyArgGlyLeuProGlyProProGlyAsnA 886
814 GAGGTAGCCAGCAGCAGCCCAAGACTGATCATGAAGGCATACAGAGTAGGCTGGCG 755
886 snGlyAsnProGlyProProGlyPro-----SerGlyAlaProGlyL 900
754 ACAGTGTGCTCG-----GGTCCCGAAGAGGTCTAGAGAGCAG 719
900 ysaspGlyProProGlyProAlaGlyAsnSerGlySerProGlyAsnProGlyIleAlaG 920
718 GGCTCCTAGTGGAGTGAAGCACACCTGGGCCACAGAGTCCAGCAGCCCGCCAGGAT 659
920 lyProLysGlyAspAlaGlyGlnProGlyGlyLysGlyPro-----ProGly- 935
658 GAGCAGTGCCAGCTCCAGGGCGCTGGATCCG---GGCACAGCAGCCCTGCTAGCCAGCC 602
936 --AlaGlnGlyProProGlySerProGlyProLeuGlyIleAlaGlyLeuThrGlyAlaA 955
601 GGCCCTTGGATGAGAAGAGCTCAGCAGATGCCCAAGCAGATGCCAGATGAAGG 542
955 rgGlyLeuAlaGlyProProGlyMetProGly---ProArgGlySerProGlyProGlnG 974
541 CCGCGCGCGCCATAGC-----GTCCACGCCAGTGTGCTACTGTGCTGAGCCTAGGAGCGG 488
974 lyIleLysGlyGluSerGlyLysProGlyAlaSerGlyHis-----AsnG 989
487 GACACAGA-----CCAGGCCAGCAGTGGACCAATGCCAGCACCATTGCTATGAACCTT 434
989 lyGluArgGlyProProGlyPro-----G 997
433 CTCTCTACCCCTTCCAGCAGCAGAGGGCGGCACATAGTGTGCTGCTGGCGCCAAACA 374
997 lnGlyLeuProGlyGlnProGlyThrAlaGly-----*GluProGlyA 1011
373 CACCTCCAGGCCAAAGTTAGCAGTTGACCAAGCAGAGCTGGGCTTTCCGGTGCC----- 318
1011 rgAspGlyAsnProGlySerAspGlyGlnProGlyArgAspGlySerProGlyGlyLysG 1031
317 -----GCAGAGCGGCTCACCCACAGCTCTGGACCATAGTGG--- 278
1031 lyAspArgGlyGluAsnGlySerProGlyAlaProGlyAlaProGly-HisProGlyPro 1050
277 CCAGCGGGTAGGCTCAGGGGGCGGTTCAGGCACTCCAGAACTGCTCTCTCGGCTCT 218
1051 ProGlyProValGlyProSerGlyLysSerGly-----AspArgGlyGluThrGlyPro 1068

QY 2313 AAAACCCCTTCTCTAGGTGTGTCCTAACCTAGGAGGCTAGCTGTCTAAACCCCTGAGCCCTGCGGCT 2223

[illegible]

594	Db	-----GluLysGlyAspProGlyProGlyAspHisGluAspAla	608
1875	QY	CCGAGCGTCGCGCAGACACACATATAGGCAGCTGACACTGGCTGGACATGGAG	1816
608	Db	:	
	QY	rProGlyGlyLysGlyPheProGlyProLeuGlyProProGlyLysAlaGlyProValG	628
1815	QY	CCATTAACAGGAGTGGGCCACTGGACACAGGAGGCAC	1773
628	Db		
	QY	yPro---ProGlyLeuGlyPheProGlyProProGlyGluArgGlyHisProGlyValPr	647
1772	QY	----TATCAGGATGGCAGGTCCAGGCAGATGCCCGCCGCGGAA	1731
647	Db		
	QY	oGlyHisProGlyValArgGlyProAspGlyLeuLysGlyGlnLysGlyAspThrIle	667
1730	QY	-----CCACCCCTGGCC	1720
667	Db	rCysAsnValThrTyProGlyArgHisGlyProProGlyPheAspGlyProGlyPr	687
1719	QY	TCGGTG-----GG	1712
687	Db	oLysGlyPheProGlyProGlnGlyAlaProGlyLeuSerGlySerAspGlyHisLysG	707
1711	QY	CTCACCCACCCACACACTCGGAGACATCACAGGCAGCCCGCAGAGCCGGGT	1654
707	Db		
	QY	yArgProGlyThrProGlyThr-----AlaGluIleProGlyProProGlyPh	723
1653	QY	-----GGAGGT-----GGCAGAGGCCACTG	1631
723	Db		
	QY	eArgGlyAspMetGlyAspProGlyPheGlyGlyGluLysGlySerProValGlyPr	743
1630	QY	TCCAGCACCCAGCTGTCATTAGG-----AAGGGAGCTCCAGGC	1591
743	Db		
	QY	oProGlyProProGlySerProGlyValAsnGlyGlnLysGlyIleProGlyAspProAl	763
1590	QY	-----TTAGGG-----CCTGGCAGGAAGCT	1571
763	Db		
	QY	aPheGlyHisLeuGlyProProGlyLysArgGlyLeuSerGlyValProGlyIleLys	783
1570	QY	GGTCATCAGGCTCCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATT	1521
783	Db		
	QY	lProArgGlyAspProGlyCysProGlyAlaGluGlyProAlaGlyIleProGlyPheL	803
1521	QY	-----	1521
803	Db	eGlyLeuLysGlyProLysGlyArgGluGlyHisAlaGlyPheProGlyValProGlyP	823
1521	QY	-----	1521
823	Db	roProGlyHisSerCysGluArgGlyAlaProGlyIleProGlyGlnProGlyLeuProG	843
1520	QY	---TGGCGAAGAACCTGCTCTCCGGTGTAGAGGAGGCCA-----GTGTAGGG	1469
843	Db		
	QY	lTyProGlySerProGlyAlaProGlyGlyLysGlyGlnProGlyAspValGlyProp	863
1468	QY	CAGATCTGAGGCTCAGAGGTGAACCGGTGAGGCGGCTGAAGCTGTACACAGGC	1409
863	Db		
	QY	roGlyProAlaGlyMetLys-----GlyLeuProGlyLeuProGlyArgp	878
1408	QY	CACACTGTGGCAGGCATGTGGCACCGGCAGCCACAGGAAGCTGCCACACTGGCCAA	1349
878	Db		
	QY	roGlyAla-----HisGlyPro-----GlyLeuProGlyIleProG	891
1348	QY	ATGACTGCTCAGTCCCGAATCGCTGCACCGCGGTCCA---TGACCAGAGAGAAGC	1292
891	Db		
	QY	lProPheGlyAspAspGlyLeuProGlyProProGlyProLysGlyProArgGlyLeuP	911
1291	QY	CAGGAGATGGCCACTGCAGAACGCCCGCAGGCTGCCATTCGGAAGCCTTCATCAT	1232
911	Db		
	QY	roGly-----PheProGlyPheProGlyGluArgGlyLysProG	924
1231	QY	GTCTCTCCGGGCTCGGTCCCGGCTCAGCTCTGGGCAGCGCTGGTACAGCCCTCGCC	1172
924	Db		
	QY	lValaGluGly-----CysProGlyAlaLys---GlyGluProGlyGluLysGlyMets	941

Db 54 SerLeuLeuThrProTyrValGlnLeuLeuGlyIleProHisLysTrpAlaAlaTyrIle 73
 QY 452 CTGGGATTGGTCCAGTGTGGCCCTGGTGTCTGCTCCCTCTAGGCTACCGAGTGAC 511
 Db 74 TrpLeuGlyGlyProIleSerGlyMetLeuValGlnProIleValGlyTyrTyrSerAsp 93
 QY 512 CACTGGCGTGGAGCGTATATGGCCCGCCCGCTTCACTCGGGCACTGTCTTGGGCATC 571
 Db 94 HisCysGlnSerSerPheGlyArgArgProPheIleAlaSerGlyAlaGlyCysVal 113
 QY 572 CTGCTGAGCTCTTTCTATCCCAAGGGCGGCTGGCTAGCAGGCGTG 619
 Db 114 AlaIleSerValIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 130
 QY 620 -----CTGTGCGCGGATCCCGAGCGCCCTGGAGCTGGCACTCTCATCTGGCGCTG 670
 Db 131 GlyAspAspMetSerLysThrLeuLysProArgAlaValThrValPheValIleGlyPhe 150
 QY 671 GGGCTGTGGACTTCTGTGGCAGGTGTCTTCACTCCAGCTGGAGCGCTCTCTCTGAC 730
 Db 151 TrpIleLeuAspValAlaAsnMetLeuGlnGlyProCysArgAlaLeuLeuAlaAsp 170
 QY 731 CTCTTCCCGG---GACCCGACCACTGTCCCGAGGCTACTCTGTCTATGCTCTCATGATC 787
 Db 171 LeuCysSerGlyAspThrArgMetArgSerAlaAsnAlaPheTyrSerPheMet 190
 QY 788 ACTCTTGGGGCTGCTGGGCTAC-----CTCCTG 817
 Db 191 AlaValGlyAsnIleLeuGlyTyrAlaAlaGlySerTyrAsnAsnLeuTyrLysLeuPhe 210
 QY 818 CTGGCATTTGACTGGGACACCACTGGCCCTGGCCCTTACCTGGGACCGAGGAGTGCC 877
 Db 211 ProPheSer-----LysThrHisAlaCysAspLeuTyrCysAlaAsnLeuLysSerCys 228
 QY 878 CTCTTTGGCTGCTCACCTCTCTTCTACCTGCTGCTAGCAGCCACTCTCTGCTGGCT 937
 Db 229 PheIleIleSerIleAlaLeuLeuIleIleIleThrValValAlaLeuSerValValArg 248
 QY 938 GAGGAGGAGCGGTGGGCGCCCGACGAGCAGCAAGAGGCTGTGCGGCCCTCTCTGTCG 997
 Db 249 GluAsnSer-----GlyProAspAspAlaAlaGluGluPro-----263
 QY 998 CCCCCTGCTCTCCATCCCGGCGCTGCTGCTTCCGGAACCTGGGCGGCGCTCTTCCC 1057
 Db 264 -----ProSerSerGlyLysIleProVal-----PheGlyGluLeuLeuGly 277
 QY 1058 CGGTGTACCACTGTGTGTGCGCATGCCCGCACCTGCGCGGCTCTTCTGGTGTGAG 1117
 Db 278 AlaLeuLysAspLeu-----ProArgProMetLeuLeuLeuLeuIleValThr 293
 QY 1118 CTGTGCTGAGTGGGCACTCATGCTTACAGCTGTTTACAGGATTTCTGGGGGAG 1177
 Db 294 CysLeuAsnTrpIleAlaTrpPheProPheIleLeuPheAspThrAspTrpMetGlyArg 313
 QY 1178 GGGCTGTACCAAGGCGGTGCCAGAGCTGAGCGGGGACCGGCGGCGGACACTATGAT 1237
 Db 314 GluIleTyrGlyGlyThr-----AlaGlyGlnGly-----LysLeuTyrAsp 327
 QY 1238 GAAGCGGTGTGGTGGGAGCTGGGCTGTCTTCTGCACTGCGCCATCTCCCTGGCTTC 1297
 Db 328 GlnGlyValArgAlaGlyAlaLeuGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1297
 QY 1298 TCTGTGTGATGAGCGGTGGTGGCGGATTTGGCACTCGAGCTGCTATTTGGCCAGT 1357
 Db 348 SerIleAlaValGluTyrLeuValArgGlyValGlyValIleLysIle-----LeuTrpGly 366
 QY 1358 GTGCGAGCTTTCTCCGTGGCTGGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1396
 Db 367 PheValAsnPheIleLeuAlaIleGlyLeuValMetThrValValSerLysValAla 386
 QY 1397 -----TCCACAGTGTG-----GCCGTGGTGACA 1420
 Db 387 GlnHisGlnArgGluHisSerAlaAsnGlyGlnLeuLeuProProSerAlaGlyValLys 406

QY 1421 GCTTCAGCC-----GCCCTCACCAGGTTACCTTCTCAGCCCTGCAGATCCTG 1468
 Db 407 AlaGlyAlaLeuSerLeuPheSerIleLeuGlyIleProLeuSerIleThrTyrSerIle 426
 QY 1469 CCCTACACATGGGCTCCCTTACACCGCGGAGAGAGAGGTGTTCTCTGCCCAATACCGA 1528
 Db 427 ProPheAlaLeuAlaSerIleTyrSerSerGlySerGlyAlaGlyGlnGlyLeuSerLeu 446
 QY 1529 GGGGACACTGGAGTGTAGCAGTACAGCAGCCTGATGACCACTTCTTCCAGGCGCT 1588
 Db 447 GlyValLeuAsnLeuAlaIleValValProGlnMetIleValSerValLeuAlaGlyPro 466
 QY 1589 AAGCCTGGA-----GCTCCCTTCCCTTAATGACACGCTGGGGCT 1627
 Db 467 PheAspSerLeuPheGlyGlyGlyAsnLeuProAlaPheValValGlyAlaIleSerAla 486
 QY 1628 GGAGGAGTGGCTGCTCCCA-----CCTCCACCGCGCTCTCGGGGCTCT 1675
 Db 487 AlaIleSerGlyValLeuAlaIleValLeuLeuProLysProSerLysAspAlaAlaSer 506
 QY 1676 GCCTGTGATGCTCTCC 1690
 Db 507 LysLeuSerLeuSer 511
 RESULT 8
 CGB07S
 collagen alpha 1(III) chain - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 04-Dec-1986 #sequence revision 04-Dec-1986 #text change 07-May-1999
 C:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946
 R:Fietzek, P.P.; Allmann, H.; Rautenberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979
 A:Title: The covalent structure of calf skin type III collagen. I. The amino acid seq
 A:Reference number: A02862; MUID:80026026; PMID:488906
 A:Accession: A02862
 A:Molecule type: protein
 A:Residues: 1-242 <PIE>
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979
 A:Title: The covalent structure of calf skin type III collagen. II. The amino acid se
 A:Reference number: A38001; MUID:80026027; PMID:488907
 A:Accession: A38001
 A:Molecule type: protein
 A:Residues: 243-422 <DEW1>
 R:Bentz, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979
 A:Title: The covalent structure of calf skin type III collagen. III. The amino acid s
 A:Reference number: A38002; MUID:80026028; PMID:488908
 A:Accession: A38002
 A:Molecule type: protein
 A:Residues: 423-571 <BEN>
 R:Lang, H.; Gnanville, R.W.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979
 A:Title: The covalent structure of calf skin type III collagen. IV. The amino acid se
 A:Reference number: A38003; MUID:80026029; PMID:488909
 A:Accession: A38003
 A:Molecule type: protein
 A:Residues: 572-808 <LAN>
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
 A:Title: The covalent structure of calf skin type III collagen. V. The amino acid seq
 A:Reference number: A38004; MUID:80026030; PMID:488910
 A:Accession: A38004
 A:Molecule type: protein
 A:Residues: 809-947 <DEW2>
 R:Allmann, H.; Fietzek, P.P.; Gnanville, R.W.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979
 A:Title: The covalent structure of calf skin type III collagen. VI. The amino acid se
 A:Reference number: A38005; MUID:80026031; PMID:488911
 A:Accession: A38005
 A:Molecule type: protein
 A:Residues: 948-1049 <ALL>

Qy	3051	AAACAGGGTGTGCGAGCTGTTGGGGAAGAAGTTGGGGTAGGGGAAGATTTGGGGTAGGGGA	2992
Dbb	8	LysSerGlyValAlaGlyGlyIleAlaGlyTyProGlyProAlaGlyProGly	27
Qy	2991	AATTTCGGGCAGTCCTTCATCAGCCCAAGTCCTAGACAGATAGAGGG	2943
Dbb	28	-----ProGlyPro- ProGlyThrSerGlyHisProGlyAl	40
Qy	2942	-AGTGAAGTGGGGAACACCAGGCTGGGCCAAGAGAGGGTGTTAGGGGAAGCCGTT	2884
Dbb	40	aProGlyAlaProGlyTyrglnGlyProProGlyGluProGlyGlnAlaGly---ProAl	59
Qy	2883	GAGACTTGAAGCCCCACCTCTACCTTCCTTAACACCCATAACCTTGGGTAAACAGCATTT	2824
Dbb	59	aGlyProGlyProGly	66
Qy	2823	GGAAATTATCATTTGGGATGACT	2785
Dbb	67	-----AlalegIyProSerGlyLysAspGlyGluSerGlyArgProGly-ArgProG	84
Qy	2784	GGGTAGGCATTTTGGGGGCCACCCACCCAGGAGGAAGAAGATTCTGGCAATGATCACCCC	2725
Dbb	84	lyProArgGlyPheProGlyPro-	91
Qy	2724	AATGACCAGCTATCTCAGGGACCTGATTTGTTGGGATCCCCACCCTACCCAATATTA	2665
Dbb	92	-----ProGlyMetLysGlyPro- -----AlaGlyMetProGlyPheProGlyMetL	107
Qy	2664	GACACCAACACAGA-	2629
Dbb	107	ys----GlyHisArgGlyPheAspGlyArgAsnGlyGluLysGlyGluProGlyAlaProG	126
Qy	2628	CTACTTTGTTAAATAAATAAGCTAAAATATTTAAATGGCTGTGTCTCTGTGTATGGCAACAG	2569
Dbb	126	ly-----LeuLysGlyGluAsnGlyValProGlyGluAspGlyAlap	140
Qy	2568	AAGGACCAACAGGCCACATCTGTATAAAGGTATAGAGGGGGGTGATCAGCAAAAAAGACA	2509

Db 408 -----GlyProGlySerProGlyP 415
 QY 1492 GTAGAGGAGCGCAGTGTCT----- 1473
 Db 415 roArgGlyGlnProGlyValMetGlyPheProGlyProLysGlyAsnAspGlyAlaProG 435
 QY 1472 --AGGCAGAGTCTCAGGGCTGAGAGGTGAACCGGTGAGGGCGCTGAAGCTGTCCAC 1415
 Db 435 lYlAsnGlyGlu-ArgGlyGlyProGlyGlyProGlyProGlnGlyProAlaGlyLys 454
 QY 1414 CAGGCGCACACT-----GTGGGACAGCATGTGCACCGGCGAGCCACAGGAAAGCTGC 1361
 Db 455 AsnGlyGluThrGlyProGlnGlyProGlyProGlyProGlyProGlyProGlyProGly 474
 QY 1360 CACACTGGCCAAATAGACTGCTCGAGTGCCTGAATCGCTGCACCA-----GCCGTCCAT 1307
 Db 475 AspThrGlyProGlyProGlnGlyLeuGlyLeu-----ProGlyThrSerGlyProp 494
 QY 1306 GACCAGAGA-----AGACCAGGAGATGGCGCACTGCAGGAACAGCCCGAG 1259
 Db 494 roGlyGluAsnGlyLysProGlyGluProGlyProLysGlyGluAlaGly---AlaProG 513
 QY 1258 GCTGCCCATCCGACGCTTCATCATAGTGTCTCCGGGCGCTCGGTGGCGGCTCAGCTCT 1199
 Db 513 lYlLeProGlyGlyLysGlyAspSerGlyAlaProGlyGluArgGlyPro----- 529
 QY 1198 GGGCAGCGCCTGTGTACAGCC-----CCTCGCCCGACGAATCCGTGTAAACAGCGTGAA 1145
 Db 530 -----ProGlyAlaGlyGlyProGlyProGlyProArgGly----- 540
 QY 1144 GGTATGAGTGCATCCAGCTGCAGCTCAGCCAGCAAGACCGCGCGAGGTGCGGG 1085
 Db 541 -----GlyAlaGlyProGlyProGlyProGlyProGlyLysGlyLysGly 553
 QY 1084 CATGCGGC-----AGCACAGCTGGTGCACCGCGGGAAGCAGGCGCCCGCAG 1040
 Db 553 lAaLaGlyProProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGly 573
 QY 1039 GTTCC---GGAAGCAAGCGGCGCCGCG-----ATGGACAGCA 1004
 Db 573 lYlGluArgGlyGlyProGlyGlyProGlyProLysGlyAspLysGlyGluProGlySerS 593
 QY 1003 GTGGGCGACA-----AGGAGGGCGCGACAGCCCTTCGTCTGG 965
 Db 593 erGlyValAspGlyAlaProGlyLysAspGlyProArgGlyProThr----- 608
 QY 964 CTCGGTGGGGCGCAGCGTCTCTCAGCCACAGCAGTGTGGCTGTCTACCGAGTCA 905
 Db 609 -----GlyProLleGlyProProGlyProAla-----G 618
 QY 904 GAAGATGAGGTGAGCAGGCAAGAGGACCTCTCTCTGGTGGCCAGGT---AGGGGCG 848
 Db 618 lYlGlnProGlyAspLysGlyGluSerGlyAla---ProGlyValProGlyLleAlaGlyP 637
 QY 847 CAGGCGACTGTCTCCAGTCAATGGCAGGAGGAGGTAGCCAGGCGCCCGCCCAAGACT 788
 Db 637 ro-----ArgGlyGlyProGlyGlu----- 643
 QY 787 GATCATGAGGCATACAGAGTAGGCTGGCGCAGCTGGT---CCGGGT---CCCGGAA 734
 Db 644 -----ArgGlyGluGlnGlyProProGlyProAlaGlyPheProGlyAlaProGlyG 661
 QY 733 GAGGTCAAGAGCAGGCGCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAG 674
 Db 661 lNAsnGlyCluProGlyAlaLysGlyGluArgGlyAlaProGlyGluLysGlyGluGlyG 681
 QY 673 CCCCACGCCAGGATGAGCAGTGCAGCTCCAGG-----GCCCTGGGATCCGGGCA 623
 Db 681 lYPro---ProGlyAlaAlaGlyProAlaGlyGlySerGlyProAlaGlyProProGlyP 700
 QY 622 CAGCAGCCTGTCTAGCAGCGCGCCCTTGGATGATGAAGAAGGCTCAGCAGATGCCCAA 563
 Db 700 roGlnGlyValLysGly-----GluArgGlySerProGlyGlyProG 714

QY 562 GGACAGTCCCCAGATAGAGGCGCGC-----GGCGCCCATAGCGTCCACGCCAGTGGTC 509
 Db 714 lYAla-----AlaGlyPheProGlyGlyArgGly----- 723
 QY 508 ACTGGCTAGCCTAGAGCGGGACACAGACAGCCAGCCAGCAGCACTGGACCAATGCCAGCAC 449
 Db 724 -----ProProGlyPro-----P 728
 QY 448 CATGGTCATGAATCTCTCTCTACCCCA-----CTTCCAGACAGCAGAGCGG 401
 Db 728 roGlySerAsnGlyAsnProGlyProGlySerSerGlyAlaProGlyLysAspGly- 747
 QY 400 CACATAGTGATGCTCGCGGCAACACACCTCCAGCCAAAGCTTAGCAGTTGACAGCAG 341
 Db 748 -----ProProGlyProProGlySerAsn-----GlyAlaProG 759
 QY 340 CAAGAGCTGGCTTTCGGGTGCGCAGCAGC-----GGCTCAC 302
 Db 759 lYSerProGlyLleSerGlyProLysGlyAspSerGlyProProGlyGluArgGlyAlaP 779
 QY 301 CCACAGCCTCTGGACCATAGTGGG---CCAGGCGGGTAGGGCTCAGGGGCGGTTCAGGC 245
 Db 779 roGlyProGlnGlyPro-ProGlyAlaProGlyProLeuGlyLleAlaGlyLeuThrGly 798
 QY 244 ACTCAGAACTGCTTCTGCTCTGCTCTCCAGAGCTCGGCGCTCTCTCT----- 194
 Db 799 AlaArgGlyLeuAlaGlyProProGlyMetProGlyAlaArgGlySerProGlyProGln 818
 QY 193 -----CCTTGTCTCCGCGCACTAGCAATCAG 164
 Db 819 GlyLleLysGlyGluAsnGlyLysProGlyProSerGlyGlnAsnGlyGluArgGlyPro 838
 QY 163 CCAGGCGCCCATTTCTGCCAGCCCTTTGGT-----GCCGTCCAGCTTCAG 116
 Db 839 ProGly-----ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGlu 853
 QY 115 CCCATGTCTC-----AACACCTGCTGTGTGGGCGACCTCAGTGGGCGACGCTCATCA 62
 Db 854 ProGlyArgAspGlyAsn-ProGlySerAspGlyLeuProGlyArgAspGlyAlaProG 873
 QY 61 CTCAGATCTCTGGCGCGCGCGCTGTCCACCGGAGCC 22
 Db 873 YAlaLysGlyAspArgGlyGluAsnGlySerProGlyAla 886
 RESULT 9
 CcHULS
 collagen alpha 1(I) chain precursor - human
 N:Alternate names: procollagen alpha 1(I) chain
 C:Species: Homo sapiens (man)
 C:Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text_change 31-Dec-2000
 C:Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567;
 S269; A29439; I53466; A02852; I37247
 R:D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.
 Gene 67, 105-115, 1988
 A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five
 A:Reference number: I60114; MUID:88329734; PMID:2843432
 A:Accession: I60114
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-369, 'L', 371-589 <DAL>
 R:Cross-references: GB:M20789; NID:g179593; PIDN:AA59373.1; PID:g179594
 R:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Pr
 Biochem. J. 253, 919-922, 1988
 A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of hum
 A:Reference number: S01143; MUID:89025644; PMID:3178743
 A:Accession: S01143
 A:Molecule type: mRNA
 A:Residues: 1-472 <TRO>
 A:Cross-references: EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID:g30016; GB:M36546;
 A:Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988
 R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams,
 Nature 310, 337-340, 1984

A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of
A:Reference number: A93335; MUID:84270697; PMID:6462220
A:Accession: A93335
A:Molecule type: DNA
A:Residues: 1-58, Q', 60-181 <CHU>
A:Cross-references: EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID:g35658
R:Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.
J. Biol. Chem. 262, 15151-15157, 1987
A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enh
A:Reference number: I55254; MUID:88033098; PMID:2822714
A:Accession: I55254
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <ROS>
A:Cross-references: GB:J02829; NID:g180387; PIDN:AAA51993.1; PID:g180388
R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A:Title: Regulatory elements in the first intron contribute to transcriptional control c
A:Reference number: A39943; MUID:88097389; PMID:3480516
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <BOR>
A:Cross-references: GB:J03559; NID:g180876; PIDN:AAA52052.1; PID:g553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.
J. Biol. Chem. 260, 2315-2320, 1985
A:Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
A:Reference number: I55237; MUID:85130970; PMID:2857713
A:Accession: I55237
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-34 <CH2>
A:Cross-references: GB:IM0627; NID:g180383; PIDN:AAA51992.1; PID:g553226
R:Wirtz, M.K.; Keene, D.R.; Horli, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollis
J. Biol. Chem. 265, 6312-6317, 1990
A:Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termi
one, type VII.
A:Reference number: A35233; MUID:90202908; PMID:2318855
A:Accession: A35233
A:Molecule type: protein
A:Residues: 33-52 <WIR>
A:Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R:Neil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A:Title: A base substitution in the exon of a collagen gene causes alternative splicing
A:Reference number: S09400; MUID:89356643; PMID:2767050
A:Accession: S09400
A:Molecule type: mRNA
A:Residues: 156-183 <WEI>
R:Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alpha
A:Reference number: A90567; MUID:71038625; PMID:5529814
A:Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A:Accession: B90567
A:Molecule type: protein
A:Residues: 162-198, Z', 200-201, Z', 203-206, Z', 208-209, Z', 211-228, B', 230, BB', 233, Z'
A:Experimental source: skin
A:Note: evidence for 170-allysine
R:Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, H.
Eur. J. Biochem. 192, 153-159, 1990
A:Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A:Reference number: S11372; MUID:90382436; PMID:2169412
A:Accession: S11372
A:Molecule type: protein
A:Residues: 175-187; 274-287, P', 289 <BAE>
R:Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A:Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
operative melting of intact type I collagen.
A:Reference number: I55342; MUID:92042092; PMID:1718984
A:Accession: I55342
A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA
A:Residues: 258-268; 1347-1357 <DEA>
A:Cross-references: GB:S67495; NID:g239007; PIDN:AAB20350.1; PID:g239008
A:Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg rep
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A:Title: Comparative study of glycopeptides derived from selected vertebrate collagen
A:Reference number: A92069; MUID:71001508; PMID:4319110
A:Accession: A92069
A:Molecule type: protein
A:Residues: 263-268 <WOR>
A:Experimental source: skin
A:Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R:Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A:Title: Segmental amplification of the entire helical and telopeptide regions of the
A:Reference number: S15989; MUID:90326017; PMID:2374517
A:Accession: S15989
A:Molecule type: mRNA
A:Residues: 281-302; 402-420; 823-843; 925-944; 1026-1045; 1143-1162 <LAB>
R:Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries,
Connect. Tissue Res. 29, 1-11, 1993
A:Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain
A:Reference number: I52905; MUID:93339042; PMID:8339541
A:Accession: I52905
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 342-352, C', 354-359 <WID>
A:Cross-references: GB:S64717; NID:g408195; PIDN:AAB27677.1; PID:g408196
A:Note: mutant sequence from patient with osteogenesis imperfecta
R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A:Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalph
A:Reference number: A90476; MUID:84080385; PMID:6689127
A:Accession: A90476
A:Molecule type: mRNA
A:Residues: 425-1250, X', 1252-1328, S', 1330-1390, X', 1392-1464 <BER>
A:Cross-references: GB:K01228; NID:g180391; PIDN:AAA51995.1; PID:g180392
A:Note: sequence partially completed for missing nucleotides by A29439
R:Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A:Title: Multixon deletion in an osteogenesis imperfecta variant with increased type
A:Reference number: A22161; MUID:85104934; PMID:2981843
A:Accession: A22161
A:Molecule type: DNA
A:Residues: 472-594, R', 596-607 <CH3>
A:Cross-references: GB:K03178; GB:K03179; NID:g179612; NID:g179613; PIDN:AAA51847.1;
A:Note: the authors translated the codon CGT for residue 595 as Pro
R:Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A:Title: Variable expression of osteogenesis imperfecta in a nuclear family is explai
A:Reference number: A35336; MUID:90252792; PMID:2339700
A:Accession: A35336
A:Molecule type: mRNA
A:Residues: 710-720, E', 722-737, E', 739-745 <WAL>
A:Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R:Forlino, A.; Zollezzi, F.; Valli, M.; Pignatti, P.F.; Cetia, G.; Brunelli, P.C.; Mot
Hum. Mol. Genet. 3, 2201-2206, 1994
A:Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in th
A:Reference number: I54365; MUID:95187161; PMID:7881420
A:Accession: I54365
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 746-766, S', 768-781 <FOR>
A:Cross-references: GB:I47667; NID:g1009093; PIDN:AAB59576.1; PID:g1009094
R:Chessler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A:Title: Mutations in the carboxy-terminal propeptide of the pro alpha 1(I) chain of
A:Reference number: A47426; MUID:93352646; PMID:8349697
A:Accession: A47426
A:Molecule type: mRNA
A:Residues: 1179-1276, H', 1278-1336, 1339-1387, R', 1389-1464 <CHE>

Alignment Scores:

US-09-759-143-110 (1-3410) x CGHU1S (1-1464)

— — —

QY 971 CTGCTGGCTCGGTGGGCGCCAGCG-----CTGCTCTCTCAGCCACCA 930
 Db 557 -----GlyProAlaGlyGlnAspGlyArgProGlyProGlyProPro 571
 QY 929 -----GCAGTGTGGGTG 918
 Db 572 GlyAlaArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAla----- 589
 QY 917 CTACCGAGGTGAGGAAGATGAGGTGAGCGAGCGCAAGAGGCACT-----CCTCTGGG 864
 Db 590 -----GlyGluProGlyLysAlaGlyGluArgGlyValProGlyProGly 605
 QY 863 TGCCCGAGT-----AGG 852
 Db 606 AlaValGlyProAlaGlyLysAspGlyGluAlaGlyAlaGlnGlyProGlyProAla 625
 QY 851 GGGCCAGGCGCACTGTGTCCCGAGTCAATGGCAGCGCAGGA----- 813
 Db 626 GlyProAlaGlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPheGlnGlyLeu 645
 QY 812 ---GATAGCCAGGAGCCCGCCAGACTGATCATGAGGCATAGACAGTAGCCCTGGC 756
 Db 646 ProGlyProAlaGlyProGlyGluAlaGlyLysProGlyGluGlnGlyValProGly 665
 QY 755 GACAGTGT-----CGGGTCCC-----GGAAGAGTGCAGAGCAGCGGCTCCAGTGGAG 705
 Db 666 AspLeuGlyAlaProGlyProSerGlyAlaArgGlyGluArg-----GlyPheProGlyGlu 684
 QY 704 TGAACACACCTGGCCACAGAAGTCCAGCAGCCCGCCAGGATGAGCAGTGCCTCA----- 648
 Db 685 ArgGlyValGlnGlyProGlyProAlaGlyProArgGlyAlaAsnGlyAlaProGly 704
 QY 647 -----CCTCAGGGGCTGGATCCGGGCGACAGCAGCC 615
 Db 705 AsnAspGlyAlaLysGlyAspAlaGlyAlaProGlyAla-----ProGlySerGlnGly 722
 QY 614 CTGTAGCCAGCGCGCCCTTGGATGAGAAGAGGCTCAGCAGGATGCCCAAGCAGCAGTG 555
 Db 723 AlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGly----- 740
 QY 554 CCAGATGAGGCGCGGGCGGCGCATAGCTCCAGCGCAGTGTCTCAGTGTGAGCCTA 495
 Db 741 ProLysGlyAspArgGlyAspAlaGlyProLysGlyAlaAspGly-----SerPro 757
 QY 494 GGAGCG---GGACACAGA-----CCAGGCGCCAGCAGTGGAC 462
 Db 758 GlyLysAspGlyValArgGlyLeuThrGlyProLleGlyProProGlyProAla----- 775
 QY 461 CAATGCCCA----- 453
 Db 776 GlyAlaProGlyAspLysGlyGluSerGlyProSerGlyProAlaGlyProThrGlyAla 795
 QY 452 -----GCACCATGTGATGAATCTCTCTACCCCGCAGCGGCTCAGCAGCGCGCA 399
 Db 796 ArgGlyAlaProGlyAspArgGlyGluProGlyProGlyProAlaGlyPheAlaGly 815
 QY 398 CATAGTGTGCTGCTGGCGCAACACACACCTCCAGGC-----CAAGGTTAGCA 351
 Db 816 -----ProGlyAlaAspGlyGlnProGlyAlaLys 826
 QY 350 GGTGTACACAGACAGCTGGGCTTCCGCTGCGGAGCAGCGGCTCAGCAGCCCTCT 291
 Db 827 GlyGluProGlyAspAlaGlyAlaLysGlyAspAla-----GlyProGlyProAla 844
 QY 290 GGACCATAGTGGCCA-----GGCGGTAGGCTCAGGGGCGGCTTACAGGCAC 243
 Db 845 GlyPro-AlaGlyProProGlyProLleGlyAsnValGlyAlaProGlyAlaLysGlyAla 864
 QY 242 TCCAGAAGCTGCTCGGTCTGCTGCT-----CCAGAAGCTGGGCGCTCTCCCTCTTG 189
 Db 864 a-----ArgGlySerAlaGlyProProGlyAlaThrGlyPheProGlyAla 879

QY 188 CTCCGCCCAACTGCCTAGGAATCAGCCAGCGCCCATTTCTGCCAGCCCTTTGGTGGCGG 129
 Db 879 aAlaGlyArgValGlyProProGlyProSerGlyAsnAlaGlyProProGlyProProGly 899
 QY 128 TCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGG-----GCACCTCAGTGGG 75
 Db 899 yProAlaGlyLysGlu-----GlyGlyLysGlyProArg-GlyG 912
 QY 74 ACAGTCTCATCACTCACTCAGATCTCTGCGCGA-----G 45
 Db 912 luthrGlyProAlaGlyArgProGlyGluValGlyProProGlyProProGlyProAlaG 932
 QY 44 GCGCGCGGCTGTCAACCGGAGCC 22
 Db 932 lyGluLysGlySerProGlyAla 939

RESULT 10
 CGCHIS
 collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
 C:Species: Gallus gallus (chicken)
 C:Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 31-Mar-2000
 C:Accession: A90458; A90181; A02857
 R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, Biochemistry 21, 2048-2055, 1982
 A:Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete pr
 A:Reference number: A90458; MUID:82231995; PMID:7093229
 A:Accession: A90458
 A:Molecule type: protein
 A:Residues: 1-1036 <HIG>
 A:Experimental source: skin
 A:Note: this is the latest in a series of papers from these workers elucidating the s
 R:Eyre, D.R.; Glimcher, M.J.
 Biochem. Biophys. Res. Commun. 48, 720-726, 1972
 A:Title: Evidence for a previously undetected sequence at the carboxyterminus of the
 A:Reference number: A90181; MUID:72243016; PMID:5047697
 A:Accession: A90181
 A:Molecule type: protein
 A:Residues: 1037-1042 <EYR>
 A:Experimental source: skin
 A:Note: residues 1037-1042 above correspond to the carboxyl end of the protein
 C:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in s
 C:Comment: Most of the prolines at the third position of the tripeptide repeating uni
 C:Comment: pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trime
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Alignment Scores:
 Pred. No.: 4,1e-13 Length: 1042
 Score: 342.00 Matches: 270
 Percent Similarity: 32.15% Conservative: 48
 Best Local Similarity: 27.30% Mismatches: 385
 Query Match: 5.49% Indels: 286
 DB: 1 Gaps: 52

US-09-759-143-110 (1-3410) x CGCHIS (1-1042)

QY 2578 ATGCCACAGAGGAGCAACAGCCCATCTGTATAAAGGTAAAGGGGGTGGATCAG 2519
 Db 55 MetGlyProArgGlyProAlaGly--ProProGlyLysAsnGlyAspGlyGluAlaG 74
 QY 2518 CAAAGACACAGTGTGGTGGCTGAGGGACCT-----GGTCTTGTGTGTGTC 2471
 Db 74 lyLysProGlyArgProGlyGlnArgGlyProProGlyProGlnGlyAlaArgGlyLeuP 94
 QY 2470 CC---CTCAGGACTCTTCCCTACAAAATAAGTCATATG-----T 2435
 Db 94 roGlyThrAlaGlyLeuProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyA 114
 QY 2434 TCAATCCCATGGAGGAGTGTTCCTCATCTAGAAACTCCCATGCAAGAGCTACATTAACG 2375
 Db 114 laLysGlyGlnProGlyProAlaGlyProLysGlyGluProGlySerProGlyGluAsnG 134


```

Db      756 rgProGlyGluProGlyProAlaGlyProProGlyProProGlyGluLysGlySerProG 776
QY      520 AGCCAGTGGTCACTGGCTGAGCCCTAGAGCGGGGACACAGACACAGGCCCA----- 471
Db      776 lyAlaAspGly-----ProIleGlyAlaProGlyThrProGlyProGlnGlyIleA 793
QY      470 -----GCACGTGACCAAT 458
Db      793 laGlyGlnArgGlyValValGlyLeuProGlyGlnArgGlyGluArgGlyPheProGlyL 813
QY      457 GCCCAGACCATGTCATGAATCTCTCTACCCCCCTTCCAGCAGCAGAGCGGCAC 398
Db      813 euProGlyProSerGly-----GluProGlyLysGlnGlyProSerGlyAlaSerGlyG 831
QY      397 ATAGTGTAGCTCGCGGCAACACACACTCCAGGCAAGGTAGCAGGTGTGACCAACAA 338
Db      831 luArgGlyProProGlyProMetGlyProProGly-----LeuAlaGlyProProGlyG 849
QY      337 GAGCTGGGCTTCCCGTGC-----GCAGAGGCGGCTCACCCACAGCCCTCTGGACCATAGT 281
Db      849 luAlaGlyArgGlyGluGlyAlaProGlyAlaGlyAlaProGlyArgAspGlyAla-Ala 868
QY      280 GGGCCA---GGCGGTAGGCTCAGGGGCGCTTCAGGCACTCCAGAACTGCTTCGCTC 224
Db      869 GlyProLysGlyAspArgGlyGluThrGlyProAlaGlyProPro----- 883
QY      223 GGCTCTGCTCCAGAGCTGCGCCCTCTCT-----CCTGCTGCGCGCACTG 176
Db      884 ---GlyAlaProGlyAlaProGlyAlaProGlyProAlaGlyLysAsnGly 902
QY      175 CCTAGGAATCAGCCAGCGCCCTCTCTGCCAGCCCTTGGTGGCGGCTCCAGCTTCTCAG 116
Db      903 AspArgGlyGluThrGlyPro-----AlaGlyProAlaGlyPro 915
QY      115 CCC 113
Db      916 Pro 916

RESULT 11
T12198
C:Species: Vicia faba (fava bean)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 20-Jun-2000
C:Accession: T12198
R:Weber, H.; Borisjuk, L.; Helm, U.; Sauer, N.; Wobus, U.
A:Title: A role for sugar transporters during seed development: molecular characterization of a sucrose transporter from Vicia faba
A:Reference number: 217451; MUID:97355984; PMID:9712465
A:Accession: T12198
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-523 <WEB>
A:Cross-references: EMBL:293774
C:Genetics:
A:Gene: sut
C:Superfamily: common tobacco sucrose transport protein

Alignment Scores:
Pred No.: 7,67e-13 Length: 523
Score: 337,50 Matches: 134
Percent Similarity: 40,38% Conservative: 76
Best Local Similarity: 25,77% Mismatches: 185
Query Match: 5,26% Indels: 125
DB: 2 Gaps: 20

US-09-759-143-110 (1-3410) x T12198 (1-523)
QY      332 CAGCTCTGCTGGTCAACCTGCTTGGCTGGAGGTGTGTTGGCCGCGCAGGCATC 391
Db      36 LysileMetValValAlaSerIleAlaAlaGlyValGlnPheGlyTrpAlaLeuGlnLeu 55
QY      392 ACCTATCTGCGGCTCTGCTGCTGGAAGTGGGGGTAGAGGAGAGTTCATGACCATGGTG 451

```

```

Db      56 SerLeuLeuThrProTyrValGlnLeuLeuGlyIleHisHisThrTrpAlaAlaTyrIle 75
QY      452 CTGGCATTTGGTCCAGCTGGGCTGGTCTGTGTCCTCCCTCTAGCTACAGCCAGTGAC 511
Db      76 TrpLeuGlyGlyProIleSerGlyMetLeuValGlnProIleValGlyTyrHisSerAsp 95
QY      512 CACTGGCTGGAGCTATGGCCCGCGGCTTCATCTGGGCACTGCTCTGGGCTGCGGCATC 571
Db      96 ArgCysThrSerArgPheGlyArgArgProPheIleAlaAlaGlySerIleAlaVal 115
QY      572 CTGCTGAGCTCTTCTCTATCCCAAGGCGGCTGGCTAGCAGGGCTG----- 619
Db      116 AlaIleAlaValPheLeuIle-----GlyTyrAlaAlaAspLeuGlyHisSerPhe 132
QY      620 -----CTGTGCGCGGATCCAGGCGGCTGGAGCTGGCAGCTGCTATCTGCTGCTGCTG 670
Db      133 GlyAspSerLeuAspGlnLysValArgProArgAlaIleGlyIlePheValGlyPhe 152
QY      671 GGCTGCTGCACTTCTGTGGCCAGGTGTGCTTCACTCCAGCTGGAGGCTGCTCTCTGAC 730
Db      153 TrpIleLeuAspValAlaAsnAsnMetLeuGlnGlyProCysArgAlaLeuLeuGlyAsp 172
QY      731 CTCTTCCGG---GACCGGACCACTGTGCGCCAGGCTCTCTCTCTATGCTTCTATGATC 787
Db      173 LeuCysAlaGlyAsnGlnArgLysThrArgAsnAlaAsnAlaPheSerPheMet 192
QY      788 AGTCTGGGGGCTGCTGGGCTACCTCTCTGCGCCAGGCTCTCTCTCTATGCTTCTATGATC 829
Db      193 AlaValGlyAsnValLeuGlyTyrAlaAlaGlyAlaTyrSerLysLeuTyrHisValPhe 212
QY      830 -----TGGGACACCACTGCGCCCTCTGCGCCAGGAGGAGGAGTGCCTCTT 883
Db      213 ProPheThrLysThrLysAlaCysAsnValTyrCysAlaAsnLeuLysSerCysPhe 232
QY      884 GGCTGCTGCACTCTCTCTCACTGCTGCTGCGGCGCGCTG-----GCTTTCCGAACTG 1042
Db      233 -----LeuSerIleAlaLeuLeuThrValLeuAlaThrSerAlaLeuIleTyrValLys 250
QY      941 GAGCAGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 988
Db      251 GluThrAlaLeuThrProGluLysThrValValThrValThrGluAspGlyLysSerGly 270
QY      989 TCCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1042
Db      271 GlyMet-----ProCysPheGlyGlnLeuSerGlyAlaPheLysGluLeu 285
QY      1043 GGCCTGCTGCTTCCCGGCTGCACAGCTGTGCTGCGCATGCGGCGGCGGCGGCGGCGGCGG 1102
Db      286 -----LysArgProMetTrpIle 291
QY      1103 CTCCTGCTGGCTGAGCTGTGCGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGT 1162
Db      292 LeuLeuLeuValThrCysLeuAsnTrpIleAlaTrpPheProPheLeuLeuPheAspThr 311
QY      1163 GATTTCGTGGCGAGGGGCTGTACAGGGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1222
Db      312 AspTrpMetGlyLysGluValThr-----GlyGlyThrValGly 324
QY      1223 CGGAGACAC---TATGATGAAGGCTTCGGATGGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 1279
Db      325 GluGlyHisAlaTyrAspMetGlyValArgGluGlyAlaLeuGlyLeuMetLeuAsnSer 344
QY      1280 GCCATCTCCCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1339
Db      345 ValValLeuGlyAlaThrSerLeuGlyValAspIleLeuAlaArgGlyValGly---Gly 363
QY      1340 GCAGCTCTATTGGCCAGTGTGGCAGCTTTCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1399
Db      364 ValLysArgLeuTrpGlyIleValAsnPheLeuLeu-----AlaIleCysLeuGly 380
QY      1400 CACAGTGGCGGCTGGT----- 1417
Db      381 LeuThrValLeuValThrLysLeuAlaGlnHisSerArgGlnTyrAlaProGlyThrGly 400

```


QY	1298	AGAAGACCAGGGAGATGGCGCACTGCAGGAACAGCC-----	1263
DB	422	GlyAsnProGlyAlaLysGlyThrProGlyGluProGlyLysAsnGlyAlaLysGlyAsp	441
QY	1262	CGAGGTGCCCATCCCGAAGCCCTTCATCATAGTGTCTCCGGGCTCGTGCCCGGCTCAG	1203
DB	442	ProGlyProLysGlyGluArgGlyGluAsnGlyThrProGlyAlaArgGlyProProGly	461
QY	1202	CTCTGGGCACGCCCTGTGTACAGGCCCTCGCCACCAAAATCCGTGTAACACAGCTGAAGG	1143
DB	462	GluGluGlyLysArgGlyAlaAsnGlyGluProGly-----Gln	474
QY	1142	TCATGATGCCATCCAGCTGCACAGCTCAGCCAGGAAGAGCGCGCA-----	1095
DB	475	AsnGlyValProGlyThr-----ProGlyGluArgGlySerProGlyPheArg	490
QY	1094	-----GGTGTCCGGGCATCGGCAGCACAGCTGGTCAGCCGGGAAGCAGGCGC	1044
DB	491	GlyLeuProGlySerAsnGlyLeuProGlyGluLysGlyProAlaGlyGluArgGlySer	510
QY	1043	CGAGTTCCTCCGAAGCCAGCGGCCCGCATGACAGCAGTGGGCGCACAAGAGGGGG	984
DB	511	ProGlyProProGlyProSerGlyProAlaGlyAsp-----ArgGly	524
QY	983	CCGACAGCCCTTCGTGTGGTCTGGTGGGGCCACGCGCTGCT-----	942
DB	525	GlnAspGly-----GlyProGlyLeuProGlyMetArgGlyLeuPro	538
QY	941	-----CCTCAGCCACCAGCA-----GTGTGGCTG	918
DB	539	GlyIleProGlySerProGlySerAspGlyLysProGlyProGlyAsnGlnGlyGlu	558
QY	917	CTAGCAGGTGAGGAAGATGAGGTGAGCAGGCCCAAGAGCAGCTCCT-----	870
DB	559	ProGlyArgSerGlyProProGlyProAlaGlyProArgGlyGlnProGlyValMetGly	578
QY	869	-----CCTGGGTGCCAGGT-----AGGGGCCAGGG	843
DB	579	PheProGlyProLysGlyAsnGluGlyAlaProGlyLysAsnGlyGluArgGlyProGly	598
QY	842	CACGTGGTGTCCAGTCAATGCAGGCAGGA-----GGTAGCCAGGCAGC	798
DB	599	GlyProProGlyThrProGlyProAlaGlyLysAsnGlyAspValGlyLeuProGlyPro	618
QY	797	CCCCAAGACTCATGAAGCATAGCAGTAGTGGCTCGCGACAGT-----	750
DB	619	ProGlyProAlaGlyProAlaGlyAspArgGlyGluProGlyProSerGlySerProGly	638
QY	749	-----GGTCCGGTCCCGAAGAGTCAAGACGA-----	720
DB	639	LeuGlnGlyLeuProGlyGlyProGlyProAla-----GlyGluAsnGlyLysProGlyGlu	657
QY	719	-----GGGCTCCAGTCAGGTAAGC-----	699
DB	658	ProGlyProLysGlyAspIleGlyProGlyPheProGlyProLysGlyGluAsnGly	677
QY	698	ACACCTGGCCACAGAAGTCCAGCAGCCCA-----CGCCCA	663
DB	678	IleProGlyGluArgGlyProGlnGlyProProGlyProThrGlyAlaArgGlyGlyPro	697
QY	662	GGATGAGCAGTGCACGT-----CCAGGGGCC	636
DB	698	Gly-----ProAlaGlySerGluGlyAlaLysGlyProProGlyProProGlyAla	714
QY	635	TGGGATCCGGGCACAGCAGCCCTGTAGCCAGCCCGGCTTGGGTGAGAAAGAGCTCA	576
DB	715	-----ProGlyGlyThrGlyLeuProGlyLeuGlnGlyMetProGlyGluArgGlyAla	732
QY	575	GCAGGATCCCCAAGACAGTGCCAGATGAAGGGCCGCG-----GGGCGCATAGCGTC	522
DB	733	SerGlySerProGly-----ProLysGlyAspLysGlyGluProGlyGlyLysGlyAla	750

521 CAC-----GCCAGTGGTCACTGGCTAGCCTAGGAGCGGGACACAGACCAGGC 474

751 AspGlyLeuProGlyAlaArgGlyGluArgGlyAsnValGlyProIleGlyProGly 770

473 CCAGCACTGCACCAATGCCAGCACCATGGTCATGAACCTTCCTCCTTA---CCCCCACTT 417

771 ProAla-----GlyProProGlyAspLysGlyGluThrGlyProAlaGlyAlaProGly 788

416 CCAGCAGCAGAGCGGCACATAGGTGATCCCTGCGGCCAACAACACACTCCAGGCCAAAGG 357

789 ProAlaGlySerArgGly-----GlyProGlyGluArg 799

356 TTAGCAGGTTGACCAACAAGAGCTGGGCTTCGGTGGCGCAGCAGGCGGCTCACCCACA 297

800 -----GlyGluGlnGlyLeuProGlyProAla-----GlyPheProGly 812

296 GCCTCTGGACCATPAGTGG---CCAGCGGGTAGGCTCAGGGGGCGGTTCAGGCACCTCC 240

813 AlaProGly-GlnAsnGlyGluProGlyGlyLysGlyGluArgGlyProGly----- 830

239 AGAACTGCTTCGCTCGGCTCTCCAGAGCTGCGGCCCTCTCCTCCTGTGTCGCCCA 180

831 -----LeuArgGlyGluAlaGlyProProGlyAlaAlaGly----- 842

179 ACTGCTAGGACATCAGCGGCCCATTTCTGCCAGCCCTTTGGTGGCGGTCCAGCTTC 120

843 ---ProGlnGlyProGlyAla-----ProGlyProPro--- 853

119 TCAGCCCATCTCAACACCTGCTGCTGGGGCACCTCAGTGGGGACAGCTCTCATCACT 60

854 -----GlyProGln-GlyValLysGlyGluArgG 863

59 CAGATCCT---GGCGAGGCGCGGCTGTCACCCGGAGCC 22

863 lySerProGlyGlyProGlyAlaAlaGlyPheProGlyAla 876

RESULT 13

I48103

type VII collagen - Chinese hamster (fragment)

C:Species: Cricetulus griseus (Chinese hamster)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Sep-1999

C:Accession: I48103

R:Greenspan, D.S.

Hum. Mol. Genet. 2, 273-278, 1993

A:Title: The carboxyl-terminal half of type VII collagen, including the non-collageno

A:Reference number: I48103; MUID:93271985; PMID:8499916

A:Accession: I48103

A:Molecule type: mRNA

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Residues: 1-1549 <RES>

A:Cross-references: GB:I06863; NID:g388624; PIDN:AAA36968.1; PID:g388625

C:Superfamily: unassigned collagens; animal Kunitz-type proteinase inhibitor homology

F:1484-1536/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

Alignment Scores:

Pred. No.:	1-01e-12	Length:	1549
Score:	335.50	Matches:	324
Percent Similarity:	30.87%	Conservative:	61
Best Local Similarity:	25.98%	Mismatches:	419
Query Match:	5.39%	Indels:	444
DB:	2	Gaps:	73

US-09-759-143-110 (1-3410) x I48103 (1-1549)

QY 3039 GGAGCTGGTGGGAAGTTGGGGGTAGGGAAGTTGGGGGTAGGGGAATAATTTTGGGCAG 2980

DB 36 GlySerProGlyProGlnGlyProAlaGlyArgAlaGlyGluLysGlyGluLysGlyAsp 55

QY 2979 TGCCTTCAT-----CAGCCC----- 2965

DB 56 CysGluAspGlyAlaProGlyLeuProGlyGlnProGlyAlaProGlyGluProGlyLeu 75

QY 2964 -----AGTCCTAGAGAGAGTACAGGGGAGTGCAGACTGGGGGAA 2926

```

Db      76 ArgGlyThrProGlyIleThrGlyProLysGlyAspArgGlyGln-----ThrGlyThr 93
QY      2925 CCAGGCTGGCCCAAGAGAGAGAGGGTGGTTAGGAAGCGGTGAGACCTGAAGCCGCCACC 2866
Db      94 ProGly-GluProGlyGluLysGlyGluArgGlyProPro-GlyProValGlyProGlnG 113
QY      2865 CTCTACCTCTCTTCAACACCCTAACCTTGGGTAAACAGCATTTGGAATTATCATTTGGGAT 2806
Db      113 LysLeuProGly-----ValAlaGly-----HisProGly- 122
QY      2805 GAGTAGAATTTCCAGGTCCTCGGTTAGCATTTTGGGGGGCCAGACCCAGAGAGAA 2746
Db      123 --ValGluGlyProGlu-----GlyProProGlyProAlaGlyArgA 136
QY      2745 GATTCTGGCAATGATCAGCCCAATGACCAAGTACAGCTATCTCAGGGACCTGATTGTGGGATC 2686
Db      136 rgGlyGluLysGlyGluProGlyArgPro-----GlyAspP 148
QY      2685 CCCCACCCTACCCA---AATATTAGACACCAACAGAAAAGCTAGCAATGATTCCCTT 2629
Db      148 roAlaValGlyProGlyGlyAlaGlyAlaLysGlyLysGlyAspAlaGlyLeuPro- 167
QY      2628 CTACTTTGTAATAATAAAGTTAAATATTAAATGCCTGTCTGTGTATGGAACAG 2569
Db      167 ----- 167
QY      2568 AAGGACCAACAGGCACATCCTGATAAAAGGTAAGAGGGGGTGGTATCAGCAAAAAGACA 2509
Db      168 --GlyProArgGlyAlaAlaGlyIleLysGlyGluGlnGly----- 180
QY      2508 GTGCTGTGGCTGAGGGACCTGGTCTTGTGTGTGTGTCCTCCCTCAGGACTCTTCCCTTACA 2449
Db      181 -----ProProGlyLeu----- 184
QY      2448 AATAAGTCATATTTCAATCCATCCATGGAGGAGTGTTCATCTCGTAAACTCCCATGCAAG 2389
Db      185 -----AlaLeuProGlyAspProGlyProLys----- 193
QY      2388 AGCTACATTAAACGAAGTGCAGGTTAAGGGCTTAGAGATGGGAAACACAGGTACTGAG 2329
Db      194 -----GlyAspProGlyAsp----- 198
QY      2328 TTTATTACGCTCCCAAAACCCCTTCTCTAGTGTGTCTCAACTAGGAGGCTAGCTGTAA 2269
Db      199 -----ArgGlyPro-----IleGlyLeuThrGlyArgAlaGly-----P 210
QY      2268 CCCTGAGCTGGGTAATCCACCTGCAGAGTCCCGCATTCACAGTCCATGGAGCCCTTCTG 2209
Db      210 roThrGlyAspSerGlyProProGlyGluLysGlyAspProGlyArgProGlyProProG 230
QY      2208 GCCTCCCTGTATAAGTCCAG-----ACTGAAACCCCTTGAAGGCC 2167
Db      230 ly---ProValGlySerArgGlyArgAspGlyGluValGlyLysGlyValGluGlyA 249
QY      2166 TC-----CAGTCAGCAGCCCTAGAGACTGGGAGAGAGG-----AGAGGAGCGCC- 2120
Db      249 snProGlyAspProGlyLeuProGlyLysAlaGlyGluArgGlyLeuArgGlyAlaProG 269
QY      2119 -----CAGCCCCCAGCTGTGCAGCTACCCACCTCAGCAGCAGGTTGGCAGCAGAGAGC 2065
Db      269 lyValArgGlyProAla-----GlyLysGlyAspGlnGlyAspP 283
QY      2064 CACATTACTTTGGAGCAACAAGAACTGGCGGCCAGCCCGGAGCCCATGGGCTAAACA 2005
Db      283 roGlyLysGlyAspArgGlySerProGlyProSer---GlyProLysGlyAspArg- 301
QY      2004 GGAGCGGGAGCTGGAGCCAGTAGGAGGCCCTCCACCCCAATGTGCTGGAAGTTTC 1945
Db      302 -----GlyLysGlyPro-----ProGlyProProGlyArgLeuValGlyAlaGlyI 318
QY      1944 TACGCTGAGTATTGTCGCAAGTCGCTCTGTGCAAACTACTACTCTGTAGCAAAATAATG 1885

```

```

Db      318 leGly-----SerArgAspLysG 324
QY      1884 GCGACCA-----GACCCA----- 1872
Db      324 LysGluProGlyGlnGluGlyProArgGlyProLysGlyAspProGlyProGlyAlaAs 344
QY      1871 -----GGCCTGGCGCAGACACCATATAGCAGTACAGACTGGCT 1832
Db      344 erGlyGluArgGlyIleGluGlyLeuArgGlyProProGlyProGlnGlyAspProGly 364
QY      1831 GAGCTGGGCAATGGAGCCCATAAACAGAGATGGGCCACCTGGGACACGACAGAGGCACT 1772
Db      364 alArgGlyProAlaGlyAspLysGlyAspArgGlyProProGlyLeuAspGlyArgAsn- 383
QY      1771 ATCCAGGATGGCGAGTCCAGGACAGATGCCCGGCCGGAACACCCCTGGCTCGGTGG 1712
Db      384 -----GlyValAspGlyLysProGlyAlaProGlyPro----- 394
QY      1711 CTCACCCACCACACACGTACGGAGACATCAGAGCAGAGAGCCCGCAGAGCGGGGTGG 1652
Db      395 -----ProGlyProHis--GlyAlaSerGlyLysAlaGlyAspProGlyArgAspGlyLe 412
QY      1651 AGGTGGGAGCAGG-----CCTCCAGCAGCCAGCTGTCCATAGG 1607
Db      412 uProGlyLeuArgGlyGluHisGlyProAlaGlyProProGlyProGlyValProG 432
QY      1606 GAAG-----GGAGCTCCAGGCTTAGG----- 1585
Db      432 YLysThrGlyGluAspGlyLysProGlyLeuAsnGlyLysAsnGlyGluProGlyAspPr 452
QY      1584 -----CCTGGCAGAAAGCTGGTTCATCAGGCTGTCTCAGCTAGCACC 1541
Db      452 oGlyLysAspGlyArgLysGlyGluLys-GlyAspSerGlyAlaProGlyArgGluGlyP 472
QY      1540 TCCAGTGTCCCTCGGTATTGGGAGAGAACACCTGTCTTCCCGGTGGTAGAGGAGGC 1481
Db      472 roAspGlyProLysGly---GluArgGlyAlaProGlyAspProGlyLeuArgGlyProp 491
QY      1480 CA-----GTGTGTAGGACAGATCTCAGGCGGTGAGAAAGTGAACCGGTGAG 1433
Db      491 roGlyLeuProGlyGln-ValGlyProGlyGlnGlyPheProGlyValProGlyAsn 510
QY      1432 GCGCGCTGAAGCTCTCACCAGCCACACTGTGGGACA----- 1395
Db      511 ThrGlyProLysGlyAspArgGlyGluThr--GlySerLysGlyGluGlnGlyLeuProG 530
QY      1394 -----GGCATGTGCACCGCAGCCACAGGAGAAAGCTGCCACACTGGCCAAATAGAC 1343
Db      530 lyGluArgGlyLeu-----ArgGlyGluProGlySerLeuProAsn---AlaGluArgL 547
QY      1342 TGCTCGAG-----TGCGAATCGCTGCA----- 1320
Db      547 euLeuGluThrAlaGlyIleLysValSerAlaLeuArgAspIleValGluThrTrpGly 567
QY      1319 --CCAGCCGGT-----CCATGACAGAGAGAACAGCAGGAGATGGCGCACTGCAG 1271
Db      567 luSerSerGlySerPheLeuProValProGluArgArgProGlyProLysGlyAspProG 587
QY      1270 GAACA-----GCCAGGCTGCCATCCGAACGCTTCATC 1235
Db      587 lyGluArgGlyProProGlyLysGluGlySerIleGlyPheProGlyGluArg----- 604
QY      1234 ATAGTGTCTCCGGCCTCGGTCCCGGCTCAGCTCTGGGGACCCCTGTAGACGCCCTC 1175
Db      605 --GlyLeuLysGlyAspArgGlyAspProGlyProGlnGlyProProGlyLeuAlaLeu 624
QY      1174 GCCCAGAAATCCGTGTAAACACAGCGTGAAGGTTCATGAGTGCATCCAGCTGCACAGCTC 1115
Db      624 lyGluArgGlyProProGly-----ProProGlyLeuAlaGlyG 637
QY      1114 AGCCAGAGAGCGCGCAGGTGGGGCATCGCGCAGCAGCAGCTGTGCGAGCGGG 1055
Db      637 luProGlyLysProGlyIlePro---GlyLeuProGly---ArgAlaGlySerAlaGlyG 655

```


Db 529 -----PheArgPheGly-----GlyGlyGly 535
 QY 2774 TTTTGGGGGCCAG-----CCCGAGGAGAAGATTCTGGCAATGATC 2730
 Db 536 AspAlaGlySerLysGlyProAsnValSerAlaGlnGluSerGlnAlaGlnAlaIleLeu 555
 QY 2729 AGCCCAATGACAGCTATCTCAGGGGA-----CCT 2700
 Db 556 GlnGlnAlaArgLeuAlaLeuArgGlyProAlaGlyProMetGlyLeuThrGlyArgPro 575
 QY 2699 GATTTGTTGGGATCCCGCCCTACCCCAATATATAGACACCAACACAGAAAGCTAGCAA 2640
 Db 576 GlyProMetGlyProProGlySerGlyGlyLeuLysGlyGluProGlyAspMetGlyPr 595
 QY 2639 TGGATTCCCTTCTACTTTGTTAAATAAATAAGTTAAATATTTAAATGCCTGTCTCTCTGT 2580
 Db 595 oGlnGlyProArg-----GlyVa 601
 QY 2579 GATGGCCNACAGAAGGACCAACAGCCACATCTCTGATAAAGTAAAGAGGGGGTGGATCA 2520
 Db 601 lGlnGlyProProGlyProThrGlyLysProGlyArgGlyArgAlaGlySerAsp-- 620
 QY 2519 GCAAAAAGACAGCTGCTGGGCTGAGGGACCTGTTCTGTGTGTGTGCTGCTGCTGCTGCTG 2460
 Db 620 ----- 620
 QY 2459 CTTCCCTCACAAATAAGTCATATATGTTCAAATCCCATGGAGAGTGTTCATCTAGAAAC 2400
 Db 620 ----- 620
 QY 2399 TCCCATGCAAGAGCTACATTAAACGAAGCTGCAGGTTAAGGGCTTTAGAGATGGGAACC 2340
 Db 621 -----GlyAlaArg-----GlyMetPr 626
 QY 2339 AGGTGACTGAGTTTATTACAGTCCCAAAACCCTTCTCTAGGTGTCTCTCACTAGGAGG 2280
 Db 626 oGlyGln-----ThrGlyProLysGly---AspArgGlyPheAspGlyLeuAlaGln 642
 QY 2279 CTAGCTGTTAAACCTGAGCTGGGTAAATCCACCTGCAGAGTCCCGCATCTCCAGTGCATG 2220
 Db 642 Y-----LeuProGlyGluLysGlyHisArgGlyAspPr 653
 QY 2219 GACCCCTTCGGCTCCCTGTATATAGTCCAGACTGAAACCCCTTGGAAAGCCTCCAGTC 2160
 Db 653 oGlyProSerGlyProPro----- 659
 QY 2159 AGGCAGCCCTAGAGACTGGGAGAGAG-----AGAGGGAC 2124
 Db 660 -GlyLeuProGlyAspGlyGluArgGlyAspGlyGluValGlyProArgGlyLe 679
 QY 2123 GCCCCAGCCCCAGCTGTGCAG-----CTACGCACCTCAGCAGCAGAGGTGGCAGCAG 2070
 Db 679 uProGlyGluProGlyProArgGlyLeuLeuGlyProLysGlyProProGlyProGly 699
 QY 2069 AGAGCCACATTAATTGGCAGCAGACAGAACTGGGGCCAGCCCGCAGCCCGCCATGGGGC 2010
 Db 699 yProPro-----Gly-ValThrGlyMetAspGlyGlnProGly---ProLysGlyVa 715
 QY 2009 TAACAGGA---GCGGGGAGCTGGGACCCAGTGGGAGGCCCTCCACCCCAATGTGCTGG 1953
 Db 715 snValGlyProGlnGlyGluProGlyProProGlyGlnGlnGlyAsnProGly----- 732
 QY 1952 AGTTTTTACCGCTGAGTATTTGGCAAGTCGCTCTTGTCAAATACTACTCTGTAGCAA 1893
 Db 733 -----AlaGlnGlyLeu-ProGlyProGlnGlyAlaIleGlyProProGlyGlu 748
 QY 1892 AGTAATGGCGGACCAACCCAGCTCGCGGCAGACACACATATAGGAGTGCAGACTGSC 1833
 Db 749 LysGlyProLeuGlyLysProGlyLeuProGlyMetProGlyValAlaAspGlyProProGly 768
 QY 1832 TGAGCTGGAGATGGAGCCCAATAACAGGATGGGCCACCTGGG---ACACAGGAAGG 1776
 Db 769 HisProGlyLys-----GluGlyProProGlyGluLysGlyGln 782

QY 1775 CACTATCCAGGATGCGGAGGTCCAGCAGATGCCCCCGCCCGGAA----- 1731
 Db 783 GlyProProGlyProGlnGlyProIleGlyTyrProGlyProArgGlyValLysGlyAla 802
 QY 1731 ----- 1731
 Db 803 AspGlyIleArgGlyLeuLysGlyThrLysGlyGluLysGlyGluAspGlyPheProGly 822
 QY 1730 -----CCACCTGGCCTCGG 1716
 Db 823 PheLysGlyAspMetGlyIleLysGlyAspArgGlyGluIleGlyProProGlyProArg 842
 QY 1715 TGGGCTCACCCACCCACACACAGTACGGAGACATCACAGCAGAGCCCGCAGACCGCGG 1656
 Db 843 --GlyGlu-----AspGlyProGluGlyProLysGlyArgG 854
 QY 1655 GTGAGGTGGGAGCAGGCGCTCCCTCCAGCACCCACCTGTCCATTAGGGAAG---GGAG 1599
 Db 854 LyGlyProAsnGlyAspProGlyProLeuGlyProThrGlyLysGlyLysLeuGlyV 874
 QY 1598 CTCAGGCTTAGGG-----CCTGGCAGGAAGCTGGTCAATCAGGCTGTCTCACTGCTAG 1545
 Db 874 alProGlyLeuProGlyTyrProGlyArgGln-----G 885
 QY 1544 CACCTCCAGTGTCCCTCGGTATT-----TGGCAGGAACACCTGTCTTCCCGGTGT 1491
 Db 885 LyProLysGlySer-IleGlyPheProGlyPheProGlyAlaAsnGlyGluLysGlyGly 904
 QY 1490 AGAGGAGCCGAGTGTGTAGGCGAGGATCTCGAGGCTGAGAAGGTGAACCCGCTGAGGG 1431
 Db 905 ArgGlyThrPro-----GlyLysProGly----- 912
 QY 1430 CGGTGAAGTGTTCACCAGGCCACACTGTGGGACAGCATGTGGCACCAGCCAGCCACAG 1371
 Db 913 -----ProArg 914
 QY 1370 GGAAGAGTCCACACTGGCCAAATAGACTGCTCGAGTCCCGAATCGCTGCACACCGCGT 1311
 Db 915 GlyGln-----ArgGlyProThrGly 921
 QY 1310 CCATGACACAGAGAGA-----AGACCAGG---AGATGGCGCACT 1275
 Db 922 ProArgGlyGluArgGlyProArgGlyIleThrGlyLysProGlyProLysGlyAsnSer 941
 QY 1274 GCAGAAACAGCCCA-----GGCTGCCATCCGAAGCCTTCATCATAGTGTCTCGGGCCT 1218
 Db 942 GlyGlyAspGlyProAlaGlyProProGlyGluArg-----GlyPro 955
 QY 1217 CGGTCCCGGCTCAGCTCTGGGCGAGCCCTGTACAGCCCTCGCCAGCAAAATCCGTGT 1158
 Db 956 AsnGlyProGlnGlyProThrGlyPheProGlyProLysGlyProProGlyProGly 975
 QY 1157 AAAACAGCGTGAAGTTCATAGTGCCA-----LeuProGlyHisProGlyGlnArgGlyGluThrGlyPhe 991
 Db 976 LysaspGly----- 1000
 QY 1130 TCCAGCTCAGCTCAGCCACAGAGAGCCCGCGGCGAGGTGCGGGGATGCGGAGCAGACA 1071
 Db 992 GlnGlyLysThrGlyProProGlyProProGlyVal---ValGlyProGlnGlyProThr 1010
 QY 1070 -----GCTGTGCGCGCGGGAAGCAGCGCCCGCTCCCGAAAGCCAGCGG--- 1020
 Db 1011 GlyGluThrGlyProMetGlyGluArgGlyHisProGlyProProGlyProProGlyGlu 1030
 QY 1019 -----CCCGCATGGACAGCAGTGGGCGCAAGAGGGGGCGCCAGCCCTCTCTG 969
 Db 1031 GlnGlyLeuProGlyValAlaGlyLysGluGlyThrLysGlyAspPro----- 1046
 QY 968 CTGGCTCGGTGGGCGCCAGCG---CTGCTCTCTAG-----CCACAGCAGGTGTGCTG 918
 Db 1047 -----GlyProAlaGlyLeuProGlyLysAspGlyProPro----- 1058

```
QY 917 CTACGAGGTGAGGAAGATGAGGTGAGCAGGCCAAGAGGAGCAGT----- 873
Db 1059 -----GlyLeuArgGlyPheProGlyAspArgGlyLeuProGlyProValGly 1074
QY 872 -----CCTCTGGTCCAGGT-----AGGGGGCCA 846
Db 1075 AlaLeuGlyLeuLysGlySerGluGlyProGlyProGlyProAlaGlySerPro 1094
QY 845 GGGCACTGGTGCCTCAGTCAATGGCAGCAGGAGTAGCCACAGCAGCCCAAGACTGA 786
Db 1095 GlyGluArgGlyProAla-----GlyAlaAlaGlyProIleGlyIlePro----- 1109
QY 785 TCATGAAGGCATAGACAGAGTAGGCTGGCAGACAGTGTCCGGTCCCGGAAGAGTCA 726
Db 1110 -----GlyArgProGlyProGlnGlyPro----- 1117
QY 725 AGAGCAGGGCTCCAGTGAAGTGAAGCAGACCTGGCCACAGAAGTCCAGAGCCCGCC 666
Db 1118 --ProGlyProAlaGlyGlyValProGlyGluLysGlyProGlnGlyPro-Al 1136
QY 665 CCAGGATGAGCAGTCCAGCTCCAGGCGCTGGGATCCGGGCACAGCAGCCCTGCTAG 606
Db 1136 aglyArgAsp-----GlyLeuGlnGlyPro--ValGlyLeuProGlyProAlaGlyPr 1153
QY 605 AGCGGGCCCTTGGGATGAG-----AAAGAGGCTCAGCAGGA 570
Db 1153 oValGlyProProGlyGluAspAspLysGlyGluIleGlyGluProGlyGlnLysG 1173
QY 569 TCCCAAGCAGACAGTCCAGATGAAGGCGCGCGCCATAGCGTCCAGCCAGTGGT 510
Db 1173 ySerLysGlyAspLysGlyGluGlnGlyProProGlyProGlyPro--GlnGlyP 1192
QY 509 CACTGGCTGAGCCT---AGGAGCGGCACACAGCAGCCAGCAGTGGACCAATGCCCA 453
Db 1192 roIleGlyGlnProGlyProSerGlyAlaAspGlyGluPro-----GlyProArgLy 1210
QY 452 GCACCATGTCTATCACTTCTCTCTACCCACACTCCAGCAGCAGAGCGGCACATAG 393
Db 1210 lnglnGlyLeu-----PheGlyGlnLysGlyAspGluGlySerArgGly----- 1224
QY 392 TGATGCTCGCGCAACACACTCCAGC---CAAAGGTTAGCAGGTTGACCAAGA 336
Db 1225 --PheProGlyProProGlyProValGlyLeuGlnGlyLeuProGlyProProGlyGlu 1243
QY 335 GCTGGCTTTCGGGT-----GCCGCA 315
Db 1244 LysGlyGluThrGlyAspValGlyGlnMetGlyProProGlyProProGlyProArgLy 1263
QY 314 GCAGGGCGCTCACCCACAGCCTCTCGACCATAGTGGCCA----- 275
Db 1264 ProSerGlyAlaProGlyAlaAspGlyPro-GlnGlyProProGlyGlyIleGlyAsnPr 1283
QY 274 -----GGCGGTGAGGCTCAGGGGCGGCTTCAGGCACCTCCAGAACTCTTCGTCT 225
Db 1283 oGlyAlaValGlyGlyLysGlyGluProGlyGluAlaGlyGluProGlyLeuProGlyG 1303
QY 224 CGGC-----TCTGCTCCAGAAC 207
Db 1303 uGlyGlyProLeuGlyProLysGlyGluArgGlyGluLysGlyGluAlaGlyProSerG 1323
QY 206 TGGGGCTCTCTCTCTGCTGGCGC-----CAACTGCTAGGAATCAGCC 162
Db 1323 yAlaAlaGlyProProGlyProLysGlyProGlyAspGlyProLysGlySerPr 1343
QY 161 AGCGGCCATTTCTGCCAGCCCTTGGT-----GCCGCTCAGCTTCAGCC 114
Db 1343 oGly-----ProValGlyPheProGlyAspProGlyProProGlyGluPr 1358
QY 113 CATGCTCAACACTGCTGTGGGACCTCAGTGGGACACGCTCATCATCATCATCATCAT 54
Db 1358 o-----Gly-ProAlaGlyGlnAspGlyProProGlyAspLysGlyAspGlyGluP 1376
QY 53 CTGGCCGAGCGCGCGGTGTCAACCGGA 25
```

Db 1376 roGlyGlnThrGly-----SerProGly 1383

RESULT 15

QJ2389
sucrose transport protein - potato
C:Species: Solanum tuberosum (potato)
C:Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 21-Jul-2000
C:Accession: JQ2389; S40310
R:Riesmeier, J.W.; Hirner, B.; Frommer, W.B.
Plant Cell 5, 1591-1598, 1993
A:Title: Potato sucrose transporter expression in minor veins indicates a role in phloem loading
A:Reference number: JQ2389; PMID:94146554; PMID:8312741
A:Accession: JQ2389
A:Molecule type: mRNA
A:Residues: 1-516 <RTE>
A:Cross-references: EMBL:X69165; NID:g439293; PIDN:CAA48915.1; PID:g439294
A:Experimental source: cv. Desiree
C:Comment: The gene encoding for this protein is highly expressed in mature leaves.
C:Superfamily: common tobacco sucrose transport protein
C:Keywords: glycoprotein; transmembrane protein
F:31-53/Domain: transmembrane #status predicted <TM1>
F:67-86/Domain: transmembrane #status predicted <TM2>
F:103-122/Domain: transmembrane #status predicted <TM3>
F:141-160/Domain: transmembrane #status predicted <TM4>
F:180-200/Domain: transmembrane #status predicted <TM5>
F:226-248/Domain: transmembrane #status predicted <TM6>
F:285-304/Domain: transmembrane #status predicted <TM7>
F:331-349/Domain: transmembrane #status predicted <TM8>
F:366-385/Domain: transmembrane #status predicted <TM9>
F:409-427/Domain: transmembrane #status predicted <TM10>
F:429-448/Domain: transmembrane #status predicted <TM11>
F:3.92/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred No.:	1-44e-12	Length:	516
Score:	333.00	Matches:	133
Percent Similarity:	41.93%	Conservative:	93
Best Local Similarity:	24.68%	Mismatches:	201
Query Match:	5.19%	Indels:	112
DB:	2	Gaps:	21

US-09-759-143-110 (1-3410) x JQ2389 (1-516)

```
QY 278 CCCACTATGTCAGAGGCTGTGGTGAGCGGCTGCTGGGACCGGAAGCCAGCTC 337
Db 24 ProLeuAlaProSerLysLeuTrp-----Lysile 33
QY 338 TTGCTGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTGGCCGAGCAGCATCAC 397
Db 34 IleValValAlaSerIleAlaAlaGlyValGlnPheGlyTrpAlaLeuGlnLeuSerLeu 53
QY 398 GTGCGGCTCTGCTGCTGGAAGTAGGGGTAGAGAGAGATTGATGACCATGGTGGGC 457
Db 54 LeuThrProTyrValGlnLeuLeuGlyIleProHisLysPheAlaSerPheIleTrpLeu 73
QY 458 ATTGGTCCAGTGTGGGCTGTGTGTCCCGCTCTAGGCTCAGCAGCAGTGG 517
Db 74 CysGlyProIleSerGlyMetIleValGlnProValValGlyTyrTyrSerAspAsn 93
QY 518 CGTGGAGCTATGGCGCGCGGCTTCATCTGGGACCTGCTTGGGACCTCTGCTG 577
Db 94 SerSerArgPheGlyArgArgProPheIleAlaAlaGlyAlaAlaLeuValMetile 113
QY 578 AGCTCTTCTTCATCCCAAGGCGGCTGGCTAGCAGGCTG----- 619
Db 114 AlaValPheLeuIle-----GlyPheAlaAlaAspLeuGlyHisAlaSerGlyAsp 130
QY 620 ---CTGTCGCCGATGCCAGCCCTGAGCTGGCAGCTGCTCATCTCTGGGCGGCTG 676
Db 131 ThrLeuGlyLysGlyPheLysProArgAlaIleAlaValPheValGlyPheTrpIle 150
QY 677 CTGGACTTCTGTCGCGAGGTGTGCTTCACTCCACTGGAGGCCCTGCTCTCTGACCTCTTC 736
```

```

Db 151 LeuAspValAlaAsnMetLeuGlnGlyProCysArgAlaLeuLeuAlaLeuSer 170
QY 737 CGGACCGGAC---CACTGTCGCGGCTACTGCTGTATGCTTCTATGATGATGCTT 793
Db 171 GlyGlyLysSerGlyArgMetArgThrAlaAsnAlaPhePheSerPheMetAlaVal 190
QY 794 GGGGGCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 832
Db 191 GlyAsnIleLeuGlyThrAlaAlaGlySerTyrSerHisLeuPheLysValPheProPhe 210
QY 833 ---GACACAGTCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCT 886
Db 211 SerLysThrLysAlaCysAspMetTyrCysAlaAsnLeuLysSerCysPhePheIleAla 230
QY 887 CTGCTCACCTCATCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 943
Db 231 IlePheIleLeuLeuSerLeuThrIleAlaLeuThrLeuValArgGluAsnGluLeu 250
QY 944 GCAGCGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1003
Db 251 ProGluLysAspGluGlnGluIleAspGluLysLeuAlaGlyAlaGly 266
QY 1004 TGTGTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 1063
Db 267 ---LysSerLysValProPhe-----PheGlyGlyIlePheGlyAlaLeu 280
QY 1064 CACAGCTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1123
Db 281 LysGluLeu-----ProArgProMetTrpIleLeuLeuLeuValThrCysLeu 296
QY 1124 AGCTGGATGGCACTCATGACCTTCACGCTGTTTACAGGATTCGTTGGCGGCGGCTG 1183
Db 297 AsnTrpIleAlaTrpPhePheLeuTyrAspThrAspTrpMetAlaLysGluVal 316
QY 1184 TACCAGGCGCTGCCAGAGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1243
Db 317 PheGlyGly-----GlnValGlyAsp-----AlaArgLeuTyrAspLeuGly 330
QY 1244 GTTCGATGGCGAGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1303
Db 331 ValArgAlaGlyAlaMetGlyLeuLeuGlnSerValLeuGlyPheMetSerLeu 350
QY 1304 GTCATGGAGCGCTGGTGGCGGATTCGCGCACTCGAGCAGTCTATTGCGGCGGCTG 1363
Db 351 GlyValGluPheLeuGlyLysIleGly---GlyAlaLysArgLeuTrpIleLeu 369
QY 1364 GCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1417
Db 370 AsnPheValLeu-----AlaIleCysLeuAlaMetThrIleLeuValThrLysMet 386
QY 1418 ---ACAGCTTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1447
Db 387 AlaGluLysSerArgGlnHisAspProAlaGlyThrLeuMetGlyProThrProGlyVal 406
QY 1448 ---TTCACGCTGCTGCGATC----- 1465
Db 407 LysIleGlyAlaLeuLeuPheAlaAlaLeuGlyIleProLeuAlaAlaThrPheSer 426
QY 1466 CTGCGCTACACACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1504
Db 427 IleProPheAlaLeuAlaSerIlePheSerSerAsnArgGlySerGlyGlnGlyLeuSer 446
QY 1505 ---CAGTGTCTGCTGCGCAATATACGAGGCGGCGGCGGCGGCGGCGGCT 1546
Db 447 LeuGlyValLeuAsnLeuAlaIleValProGlnMetLeuValSerLeuValGlyGly 466
QY 1547 ACAGTGGAGGAGCGCTGATGACAGCTTCCTGCGGCGGCGGCGGCGGCGGCTGCT 1606
Db 467 ProTrpAspAspLeuPheGlyGlyGlyAsnLeuProGly-----Phe 480
QY 1607 CCTAATGGACAGCTGGCTGGAGGCGAGTGGC-----CTGCTCCGACCT 1651

```

Db 481 ValValGlyAlaValAlaAlaAlaAlaSerAlaValLeuAlaLeuThrMetLeuProSer 500

QY 1652 CCACCCGCGCTCTCGGGGCGCTCTCGCTGTGATGCTCTCCGTACGTGGTGGTGGT 1708

Db 501 Prop-Orlaa-----AspAlaLysProAlaValAlaMetGly 512

RESULT 16

CGHUIV

collagen alpha 1(V) chain precursor - human

N:Alternate names: procollagen alpha 1(V) chain

C:Species: Homo sapiens (man)

C:Date: 22-Nov-1993 #sequence revision 03-Oct-1995 #text change 16-Jun-2000

C:Accession: S18802; S16024; A61142; S11303; S03978; S43642; S58665

R:Greenspan, D.S.; Cheng, W.; Hoffman, G.G.

J. Biol. Chem. 266, 24727-24733, 1991

A:Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution o

A:Reference number: S18802; MUID:92105142; PMID:1722213

A:Accession: S18802

A:Molecule type: mRNA

A:Residues: 1-1838 <GRE>

A:Cross-references: GB:M76729; NID:g189519; PIDN:AAA59993.1; PID:g189520

R:Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaoi, Y.; Kato, I.

J. Biol. Chem. 266, 13124-13129, 1991

A:Title: Complete primary structure of human collagen alpha-1(V) chain.

A:Reference number: S16024; MUID:91302336; PMID:2071595

A:Accession: S16024

A:Molecule type: protein

A:Residues: 1-81, 'OL', '84', '388', 'A', '391-676', 'K', '678-1294', 'PS', '1297', 'RS', '1300-1553', 'R', '15

A:Cross-references: GB:D90279; NID:g219509; PIDN:BAAL4323.1; PID:g219510

A:Note: parts of this sequence were determined by protein sequencing

R:Yaoi, Y.; Hashimoto, K.; Takahara, K.; Kato, I.

Exp. Cell Res. 194, 180-185, 1991

A:Title: Insulin binds to type V collagen with retention of mitogenic activity.

A:Reference number: A61142; MUID:91224163; PMID:1709100

A:Accession: A61142

A:Molecule type: protein

A:Residues: 823-824, 'X', '826-842' <VAO>

A:Note: the residue designated 'X' is probably glycosylated hydroxylysine; this cyano

R:Yaoi, Y.; Hashimoto, K.; Koitabashi, H.; Takahara, K.; Ito, M.; Kato, I.

Biochim. Biophys. Acta 1035, 139-145, 1990

A:Title: Primary structure of the heparin-binding site of type V collagen.

A:Reference number: S11303; MUID:90366601; PMID:2203476

A:Accession: S11303

A:Molecule type: protein

A:Residues: 823-824, 'X', '826-848', 'I', '850-851', 'P', '853', 'PR', '856-893', 'D', '895-932', 'X', '934-

A:Note: the residues designated 'X' are probably glycosylated hydroxylysine; this seq

R:Seyer, J.M.; Kang, A.H.

Arch. Biochem. Biophys. 271, 120-129, 1989

A:Title: Covalent structure of collagen: amino acid sequence of three cyanogen bromid

A:Reference number: S03978; MUID:89227189; PMID:2496661

A:Accession: S03978

A:Molecule type: protein

A:Residues: 621-640, 'G', '642-649', 'L', '651-662', 'E', '664-667', 'Q', '669-676', 'Q', '678-683', 'P', '6

<SEY>

A:Note: there are a number of inconsistencies in the sequences in figures 6 and

R:Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champlaud, M.F.; Boutilon, M.M.;

Eur. J. Biochem. 221, 987-995, 1994

A:Title: Diversity in the processing events at the N-terminus of type-V collagen.

A:Reference number: S43642; MUID:94237164; PMID:8181482

A:Accession: S43642

A:Molecule type: protein

A:Residues: 565-576, '576-758', 'X', '760-763', 'X', '765-772', '1012-1029', '1219-1232', '1465-1474', 'X'

R:Fessler, L.I.; Brosh, S.; Chapin, S.; Fessler, J.H.

J. Biol. Chem. 261, 5034-5040, 1986

A:Title: Tyrosine sulfation in precursors of collagen V.

A:Reference number: A56977; MUID:86168226; PMID:3082875

A:Contents: annotation; identification of tyrosine sulfate in the amino-terminal prop

R:Lee, S.; Greenspan, D.S.

Biochem. J. 310, 15-22, 1995

A:Title: Transcriptional promoter of the human alpha-1(V) collagen gene (COL5A1).

A:Reference number: S58665; MUID:95374437; PMID:7646438

A:Accession: S58665

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA
A:Residues: 1-36 <LEE>
A:Cross-references: GB:L38808; MID:g1020325; PIDN:AA79853.1; PID:g1020326
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C are 5-hydroxylated and subsequently O-glycosylated).
C:Comment: A long form of the mature protein containing part of the amino-terminal propeptide the heterotrimers are probably processed to the long form.
C:Genetics:
A:Gene: GDB:COL5A1
A:Cross-references: GDB:L31457; OMIM:120215
A:Map position: 9q34.2-9q34.3
C:Complex: type V collagen may be a homotrimer of alpha 1(V) chains, a heterotrimer of alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the length, is formed with desmosine cross-links made from lysine and allysine residues
C:Function:
A:Description: structural component of extracellular fibrous polymer associated with cell A:Note: may play a role in controlling the lateral growth of collagen I fibrils
C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F:1-37/Domain: signal sequence #status predicted <SIG>
F:36-261/Domain: PAPP-like #status predicted <PAPP>
F:38-341/Domain: amino-terminal propeptide #status predicted <PRO>
F:342-565/Product: collagen alpha 1(V) chain, short form #status predicted <MAT>
F:542-558/Region: amino-terminal nonhelical telopeptide
F:559-1572/Region: helical
F:645-647/Region: cell attachment (R-G-D) motif
F:663-665/Region: cell attachment (R-G-D) motif
F:897-929/Region: heparin binding
F:1573-1605/Region: carboxyl-terminal nonhelical telopeptide
F:1606-1838/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:1615-1837/Domain: fibrillar collagen carboxyl-terminal homology <FC>
F:38/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:62-244,183-237/Disulfide bonds: #status predicted
F:159,176,385,1672,1741/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:334,236,240,262,263,273,274,275,277,279,280,338,340,346,347,352,357,416,417,420,421/Modified site: allysine (Lys) #status predicted
F:541-542/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
F:542/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:570,576,621,639,648,654,657,675,678,690,693,696,705,717,720,726,732,741,750,753,756,758 site: 4-hydroxyproline (Pro) #status experimental
F:627,642,687,708,744,774,795,804,807,810,819,825,846,864,882,897/Modified site: 5-hydroxyproline
F:627,642,687,774,795,804,807,810,819,825,846,864,882,897,1482/Binding site: carboxydrate
F:827,744/Binding site: carboxydrate (Lys) (covalent) #status experimental
F:1482/Modified site: 5-hydroxylysine (Lys) #status predicted
F:1605/1606/Cleavage site: Ala-Asp (procollagen C-endopeptidase) #status predicted
F:1639,1645,1662,1671/Disulfide bonds: interchain #status predicted
F:1680-1835,1746-1789/Disulfide bonds: #status predicted

Alignment Scores:			
Pred. No.:	1.44e-12	Length:	1838
Score:	333.00	Matches:	311
Percent Similarity:	31.57%	Conservative:	59
Best Local Similarity:	26.54%	Mismatches:	381
Query Match:	5.35%	Indels:	424
DB:	1	Gaps:	68
US-09-759-143-110 (1-3410) x CGHULV (1-1838)			
QY	3217	AAACGGCACTTAACACCCCTTGACAGATAGACCTCCCTTAGCTCAG-----	3170
DB	708	LysGlyAsnValGlyProGlnGlyLeuProGlyGlnGlnGlyAsnProGly	727
QY	3169	GCAGGGGGCTCCGAGCTTCTGTGTGAGATTCCCAAGCACAGATATACTCTGGGGC	3110
DB	728	AlaGlnGlyLeuProGly-----ProGlnGlyAlaIleGlyProProGly	742
QY	3109	TGAGATGGACAAAGGCTTGGGAAACCGCACTTTGT-----GCTCTGCTCCT	3063
DB	743	GluLysGlyPro---LeuGlyLysProGlyLeuProGlyMetProGlyAlaAspGlyPro	761
QY	3062	GCAGTAGCTCCAACAGGTTGTGGAGCTGTGGGGAAGTTGGGGTAGGGAAAGTTG	3003
DB	762	ProGlyHisProGlyLvsGln--GlyProProGlyGluLysGlyGlnGlnGlyProProG	781

Db	1561 yProProGly 1564	
RESULT 17		
S23810	collagen alpha 1(XVI) chain precursor - human	
N:Alternate names: procollagen alpha 1(XVI) chain		
C:Species: Homo sapiens (man)		
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 20-Sep-1999		
C:Accession: S23810; PQ0612; S08012		
R:Pan, T.C.; Zhang, R.Z.; Mattel, M.G.; Timpl, R.; Chu, M.L.		
Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992		
A:Title: Cloning and chromosomal location of human alpha1(XVI) collagen.		
A:Reference number: S23810; MUID:92335339; PMID:1631157		
A:Accession: S23810		
A:Molecule type: mRNA		
A:Residues: 1-1603 <PAN>		
A:Cross-references: EMBL:M92642; NID:g180757; PIDN:AAA58427.1; PID:g180758		
A:Experimental source: skin fibroblasts		
R:Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamakoshi, J. Biochem. 112, 856-863, 1992		
A:Title: Molecular cloning and partial characterization of a novel collagen chain, alpha 1(XVI)		
A:Reference number: PQ0612; MUID:93203161; PMID:1284248		
A:Accession: PQ0612		
A:Molecule type: mRNA		
A:Residues: GGR, 421-536, 'P', 538-1159, 'P', 1161-1162, 'P', 1164, 'P', 1166-1603 <YAM>		
A:Cross-references: GB:SS7132; NID:9298641; PIDN:AB25797.1; PID:g298642		
A:Experimental source: placenta		
R:Kimura, S.		
submitted to the EMBL Data Library, April 1989		
A:Description: Partial nucleotide and amino acid sequence of a collagen-like protein from human placenta		
A:Reference number: S08012		
A:Accession: S08012		
A:Molecule type: mRNA		
A:Residues: 403-419, 'GR', 421-536, 'P', 538-846, 'VM' <KIM>		
A:Cross-references: EMBL:X14963; NID:g299984; PIDN:CAA33085.1; PID:g930048		
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (Gly-X-Lys) are O-glycosylated.		
C:Genetics:		
A:Gene: GDB:COL16A1		
A:Cross-references: GDB:134045; OMIM:120326		
A:Map position: lp34-1p34		
C:Complex: type XVI collagen may be a homotrimer, or a heterotrimer of two alpha 1(XVI)		
C:Function:		
A:Note: May play a role in forming elastic connections at fibril surfaces		
C:Superfamily: unassigned collagens		
C:Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; hydroxyllysine		
F:1-21/Domain: signal sequence #status predicted <SIG>		
F:22-1603/Product: collagen alpha 1(XVI) chain #status predicted <MAT>		
F:22-333/Domain: amino-terminal nonhelical #status predicted <NC11>		
F:334-1577/Region: interrupted helical		
F:334-360/Domain: collagenous COL10 #status predicted <COL10>		
F:375-505/Domain: collagenous COL9 #status predicted <COL9>		
F:521-554/Domain: collagenous COL8 #status predicted <COL8>		
F:539-541/Region: cell attachment (R-G-D) motif		
F:572-630/Domain: collagenous COL7 #status predicted <COL7>		
F:652-722/Domain: collagenous COL6 #status predicted <COL6>		
F:738-875/Domain: collagenous COL5 #status predicted <COL5>		
F:887-938/Domain: collagenous COL4 #status predicted <COL4>		
F:973-987/Domain: collagenous COL3 #status predicted <COL3>		
F:1005-1007/Region: cell attachment (R-G-D) motif		
F:1011-1432/Domain: collagenous COL2 #status predicted <COL2>		
F:1226-1228/Region: cell attachment (R-G-D) motif		
F:1472-1577/Domain: collagenous COL1 #status predicted <COL1>		
F:1578-1603/Domain: carboxyl-terminal nonhelical #status predicted <NC01>		
F:47,327/Binding site: carbohydrate (Asn) (covalent) #status predicted		
Alignment Scores:		
Pred. No.:	2,18e-12	Length: 1603
Score:	330.00	Matches: 316
Percent Similarity:	31.43%	Conservative: 81
Best Local Similarity:	25.02%	Mismatches: 417
Query Match:	5.30%	Indels: 451

QY 436 CTTCTCTCTACCCCA-----CTTCAGCAGCAGAGCGG----- 402
 Db 1264 -----ProGlySerThrGlyArgProGlyAlaGlyGluProGlyAlaM 1280
 QY 401 -----GCACATAGGTGATGCTCGCGCCCAACACAC 371
 Db 1280 etGlyProGlnGlyArgProGlyHis-----ValGlyProGlyP 1297
 QY 370 CTCAGGCCAAAGTTAGCAGTTGACCAAGAGAGTGGCTTTCCGGT----- 321
 Db 1297 roProGlyGln-----ProGlyProAlaGlyIleSerAlaValGlyLeuL 1312
 QY 320 -----GCCGAGCAGCGCGGTCTCACCCACAG 296
 Db 1312 yGlyAspArgGlyAlaThrGlyGluArgGlyLeuAlaGlyLeuProGlyGlnProGlyP 1332
 QY 295 CCGTCGGACCATAGTGGG-----CCAGCGGGTAGGCTCAGGCGCGGTCA 248
 Db 1332 roProGly-HisProGlyProGlyGluProGlyThrAspGlyAlaAlaGlyLysGlu 1351
 QY 247 GCACATCCAGAACTCTCTCGCTCGCTGCTCCAGAGCTCGCGCTCTCTCTCTG 188
 Db 1352 GlyProGlyLysGlnGlyPheTyrGlyProGlyProGlyProGlyAspProGlyAla 1371
 QY 187 TCCCGCCAACTGCCTAGGAATCAGCGCGCGCCATTCTGCCAGCGCTTTGGTGC 128
 Db 1372 AlaGlyGln-----LysGlyGlnAlaGlyGluLysGlyArgAlaGlyMetProGlyGly 1389
 QY 127 CCAGCTTCTACGCCATCCTCAACACTGCTGCTGTGGGGACCTCAGTGGGGACAGTC 68
 Db 1390 ProGlyLysSerGlySerMetGly-ProValGlyProGlyProAlaGlyGluArgG 1409
 QY 67 TCATCACTCAGATCTCGCGAGGCGCGGCTGTCACCCGGA 25
 Db 1409 yHisProGlyAlaProGlyProSerGly-----SerProGly 1421
 RESULT 18
 G84441
 probable sucrose-proton symporter [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: G84441
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: G84441
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-594 <STO>
 A:Cross-references: GB:AE002093; NID:g3461813; PIDN:AAC32907.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g02860
 A:Map position: 2
 C:Superfamily: common tobacco sucrose transport protein
 Alignment Scores:
 Pred. No.: 2,69e-12 Length: 594
 Score: 328.50 Matches: 120
 Percent Similarity: 38.63% Conservative: 72
 Best Local Similarity: 24.14% Mismatches: 174
 Query Match: 5.12% Indels: 131
 DB: 2 Gaps: 18
 US-09-759-143-110 (1-3410) x G84441 (1-594)
 QY 347 AACCTGTAACCTTTGGCTGGAGGTGTGTTGGCGCAGGCATCACCTAT----- 397
 Db 59 SerLeuValThrLeuValLeuSerCysThrValAlaAlaGlyValGlnPheGlyTrpAla 78

398 -----GTCCCGCTCTGCTGCTGGAAGTGGGGTAGAGAGAGTTCATG 442
 Db 79 LeuGlnLeuSerLeuLeuThrProTyrIleGlnThrLeuGlyIleSerHisAlaPheSer 98
 QY 443 ACATGGTCTGGCATTTGGTCCAGTGTGGCCCTGGTCTGTGCTCCGCTCTAGAGCTCA 502
 Db 99 SerPheIleThrLeuValGlnPheValGlnPheValGlnPheValGlnPheValGln 118
 QY 503 GCAGTGGACCATGGCTGATGGCCCGCCGCTTCTATCGGCACATGTC 562
 Db 119 TrpSerAspLysCysThrSerLysTyrGlyArgArgProPheIleLeuValGlySer 138
 QY 563 TTGGGCATCTGCTGAGCTCTTTTCATCCCAAGGCGCGCTGCTAGCA-----GGG 616
 Db 139 PheMetIleSerIleAlaValIleIleIle-----GlyPheSerAlaAspIleGly 155
 QY 617 CTGCTGTGCCCGAT-----CCGAGGCGCCCTGGAG 646
 Db 156 TyrLeuLeuGlyAspSerLysGluHisCysSerThrPheLysGlyThrArgThrArgAla 175
 QY 647 CTGGCATCTCATCTCGGCGCTGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 706
 Db 176 AlaValValPheIleIleGlyPheTyrLeuLeuAspLeuAlaAsnAsnThrValGlnGly 195
 QY 707 CCACTGGAGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 766
 Db 196 ProAlaArgAlaLeuLeuAlaAspLeu-----SerGlyProAspGlnArgAsnThrAlaAsn 214
 QY 767 TCTGCTATCCCTCATGATCAGTCTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 826
 Db 215 AlaValPheCysLeuTrpMetAlaIleGlyAsnIleLeuGlyPheSerAlaGlyAlaSer 234
 QY 827 -----GACTGGGACACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 874
 Db 235 GlyLysTrpGlnGluTrp-----PheProPheLeu-----ThrSerArgAla 248
 QY 875 TGCTCTCTTTGGC-----CTGCTCACCTCATCTCTCTCTCAC 910
 Db 249 CysCysAlaAlaCysGlyAsnLeuLysAlaAlaPheLeuLeuAlaValValPheLeuThr 268
 QY 911 ---TGCTAGCAGCCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 967
 Db 269 IleCysThrLeuValThrIleTyrPheAlaLysGluIleProPheThrSerAsnLysPro 288
 QY 968 GCAGAA---GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1024
 Db 289 ThrArgIleGlnAspSerAlaProLeuLeuAspAsp----- 300
 QY 1025 TTGCTTTTCCGGAACCTG----- 1042
 Db 301 LeuGlnSerLysGlyLeuGluHisSerLysLeuAsnAsnGlyThrAlaAsnGlyIleLys 320
 QY 1042 ----- 1042
 Db 321 TyrGluArgValGluArgAspThrAspGluGlnPheGlyAsnSerGluAsnGluHisGln 340
 QY 1043 -----GGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1078
 Db 341 AspGluThrTyrValAspGlyProGlySerValLeuValAsnLeuLeuThrSerLeuArg 360
 QY 1079 CGCATGCCCGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1138
 Db 361 HisLeuProProAlaMetHisSerValLeuIleValMetAlaLeuThrTrpLeuSerTrp 380
 QY 1139 ATGACCTTCAGCTGCTTTTACACGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1198
 Db 381 PheProPhePheLeuPheAspThrAspThrMetGlyArgGluValTyrHisGlyAspPro 400
 QY 1199 AGAGCTGAGCGCGGACCGAGCGCGGAGACACTATGATGAAGCGCTTCGATGGGCGAGC 1258
 Db 401 ThrGlyAsp-----SerLeuHisMetGluLeuTyrAspGlnGlyValArgGluGlyAla 418
 QY 1259 CTGGGGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1318

1818 GAGCCATAACAGGATGGGCGCA-----CCTGGGACACAGGAAG----- 1776
QY
500 aGlyProGlnGlyLysValGlyProSerGlyAlaProGlyGluAspGlyArgProGlyPr 520
Db
1775 -----CCTATCCAGGATGGCGAGG 1756
QY
520 oProGlyProGlnGlyAlaArgGlyGlnProGlyValMetGlyPheProGlyProLysGI 540
Db
1755 TCCAGGCAGATGCCCCGGC-----CCGNAACCCCTGGCCTCG 1717
QY
540 yAlaAsnGlyGluProGlyLysAlaGlyGluLysGlyLeuPro-GlyAlaProGlyLeuA 560
Db
1716 GTGGGCTACCCACACACACAGGATCGGAGACATCACAGGAGAGGCGCCCGAGGCGG 1657
QY
560 rGlyLeu-ProGlyLysAsp---GlyGluThrGlyAlaAlaGlyProProGlyProAla 578
Db
1656 GTGGAGGTGGGAGGAGG----- 1639
QY
579 GlyProAlaGlyGluArgGlyGluGlnGlyAlaProGlyProSerGlyPheGlnGlyLeu 598
Db
1638 CCAGTCCCTCCAGCAGCCACGCTGCTCCATTAGGGAAG----- 1603
QY
599 ProGlyProProGlyProProGlyGluGlyGlyLysProGlyAspGlnGlyValProGly 618
Db
1602 -----GGAGCTCCAGGCTTA---GGGCTGCGCAGGAAG-----TGCTCATC 1564
QY
619 GluAlaGlyAlaProGlyLeuValGlyPro-ArgGlyGluArgGlyPheProGlyGluAr 638
Db
1563 AGGCTGTCTACTGCTACACCTCCAGTGTCCCTCGGTATTTGGGAGGACACCTGCG 1504
QY
638 gGlySerProGlyAlaGlnGlyLeuGlnGlyAlaArgGlyLeu---ProGlyThrProGI 657
Db
1503 TTCTCCCGGTGTAGAGGAGG-----CCAGTGTCTAGGCGAGGATCTCC 1459
QY
657 yThrAspGlyProLysGlyAlaSerGlyProAlaGlyProProGlyAlaGlnGlyProPr 677
Db
1458 AGGCTGTGAGAAGGTGAACCGGTGAGGCGGTGAAGCTG-----TCACACCGGCCAC 1406
QY
677 oGlyLeu-GlnGlyMetProGlyGluArgGlyAlaAlaGlyIleAlaGlyProLysGlyA 697
Db
1405 ACTGTGGGACAGCATGTGGCACCAGCCAGCAGGGAAGCTGCCACACTGGCAATA 1346
QY
697 spArgGly---AspValGlyGluLysGlyProGluGlyAlaProGlyLysAspGlyGlyA 716
Db
1345 GACTGTCTGAGTCCGGAATCGCT-----GCACACCGCGGTCCATGACACAGAGA----- 1296
QY
716 rGlyLeuThrGlyProLysGlyAlaSerGlyProAlaGlyAlaAsnGlyGluLysGlyG 736
Db
1295 -----AGACAGGAGATGCGGCACTGCGAGAACACGCCAGGCTGCCATCCGAAC 1244
QY
736 luValGlyProProGlyProAlaGlyThrAlaGly---AlaArgGlyAlaProGlyGluA 755
Db
1243 GCCTTCATCATAGTGTCCGGGCGCTCGGTGCGCGCTCAGCTCTGGCAGCGCCCTGGTA 1184
QY
755 rg-----GlyGluThrGlyProProGlyProAlaGlyPheAlaGlyProProGlyA 772
Db
1183 CA----- 1182
QY
772 laAspGlyGlnProGlyAlaLysGlyGlnGlyGluAlaGlyGlnLysGlyAspAlaG 792
Db
1181 --GCCCTCGCCCGCAAGAAATCGTGTAAACACCGGTGAAGTTCATGATGCCATCCAGCT 1124
QY
792 lyAlaProGlyProGlnGlyPro---SerGlyAlaProGlyProGlnGlyProThrGlyG 811
Db
1123 GCAGAGCTACGACGAGGAGGCGGCGCAGGCTGCGGGCATCGGCGGACACAGCT---- 1068
QY
811 alThrGly---ProLysGlyAlaArgGlyAlaGlnGlyProProGlyAlaThrGlyPhe 830
Db
1067 --GGTGCAGCGGGGAAGCAGGCGCCAGGTTCC-----GGAAGCCAGCGGCGCCGCGCA 1013
QY
830 rGlyAlaAlaGlyArgValGlyProProGlySerAsnGlyAsnProGlyProGlyP 850
Db

1012 TG-----GACACAGTGGGCGCA 995
QY
850 roProGlyProSerGlyLysAspGlyProLysGlyAlaArgGlyAspSerGlyProProG 870
Db
994 CAAGGAGGGGCGGACAGCGCTTCTGCTGGTGGGCGCCAGGCTGCTCTCTCTCTCTCTCT 939
QY
870 lyArgAlaGlyAspProGlyLeu-----GlnGlyProAlaGlyProProGlyG 886
Db
938 -----CAGCCACAGCAGTGTGGC 920
QY
886 luLysGlyGluProGlyAspGlyProSerGlyProAspGlyProPro----- 902
Db
919 TGTACCCAGGTGAGGAAGATGAGGTGAGCAGGCCCAAGAGCAGCCTCTCTCTGGTGGCG 860
QY
903 -----GlyProGlnGlyLeuAlaGlyGlnArgGlyIle---ValGlyLeuP 917
Db
859 CAGGT-----AGGGG---CCAGGCGCAGTGTCTCCAGTCAATGCGCAGGCGAG 815
QY
917 rGlyGlnArgGlyGluArgGlyPheProGlyLeuProGlyProSer----- 932
Db
814 GAGTAGCCCGCAGGAGCCCGCAAGACTGATCATGAAGGCATACACAGAGTAGGCTCTGGCG 755
QY
933 --GlyGluProGly-----LysGlnGlyAlaProGlyA 943
Db
754 ACAGTGT-----CCGGGT-----CCCGGAAGAGGTCCAGAGAGCAG 719
QY
943 laSerGlyAspArgGlyProProGlyProValGlyProProGlyLeuThrGlyProAlaG 963
Db
718 GGCCTCCAGTGGAGTGAAGCAGCACCTGCGCCACAGAAGTCCAGCAGCCCGCCAGGAGG 660
QY
963 lyGluProGlyArgGluGlyThrProGlyAlaAspGlyPro-----ProGlyA 979
Db
659 -----TGAGCAGTCCAGCTCCAG 641
QY
979 rgAspGlyAlaAlaGlyValLysGlyAspArgGlyGluAlaGlyAlaLeuGlyAlaProG 999
Db
640 GGGCC-----TGGGATCCGGGACAGCAGCCCTGCTAGCCAGCGCGCTGG 593
QY
999 lyAlaProGlyProProGlySerProGlyProAlaGly-----ProThrG 1014
Db
592 GA-----TGAGAAGAGGCTCAGCAGGATGCCCAAGCAGACAGTCCAGATGAAGGCGCG 539
QY
1014 lyLysGlnGlyAspArgGlyGluAlaGly-AlaGlnGlyPrometGlyProAlaGlyPro 1033
Db
538 GCGGC-----GGCATAGGCGTC 522
QY
1034 AlaGlyAlaArgGlyLeuProGlyProGlnGlyProArgGlyAspLysGlyGluAlaGly 1053
Db
521 CACGCCAGTGTCTAGCTGCTGAGCCTAGGACGCGGACACAGA----- 480
QY
1054 GluAlaGlyGluArgGlyLeuLys-----GlyHlsArgGlyPheThrGlyLeuGln 1070
Db
479 -----CCAGGCGCCAGCAGTGGACCAA-----TGCCACAGCA---CCA 447
QY
1071 GlyLeuProGlyProProGlyProSerGlyAspGlnGlyAlaSerGlyProAlaGlyPro 1090
Db
446 TGGTCATGAACCTCTCTCTACCCCTCTT-----CCAGCAGCAGCAGGCGCGCATAGTGA 390
QY
1091 SerGlyProArgGlyProProGlyProValGlyProSerGlyLysAspGlyAlaAsnGly 1110
Db
389 TGCCTCGCGCCCAACACACCTCCAGGCCAAAGTTAGCAGGTTGACACAGCAGAGCTGGG 330
QY
1111 ileProGlyProIleGlyProProGlyProArg-----GlyArgSerGly 1125
Db
329 CTTTCCGGT---GCCGCGCAGCGCGCTCACCCACAGCCTCTGGACCATAGTGGGCCAGG 273
QY
1126 GluThrGlyProAlaGlyProProGlyAsnProGlyProProGlyProProGlyProPr 1145
Db
272 CGGG-----TAGGGCTCAGGGGCGGCTTCCAGGCACTCCAGA 237
QY
1145 oGlyProGlyIleAspMetSerAlaPheAlaGlyLeuGlyProArgGlyLysGlyProAs 1165
Db
236 ACTGCTTCTGCTCGGCTCTGCTCCAGAGCTCGCGGCTCTCTCTCTCTCTCTCTCTCTCT 177